Feature Based Control of Compact Disc Players

Odgaard, Peter Fogh

Publication date:
2004

Document Version
Også kaldet Forlagets PDF

Feature Based Control of Compact Disc Players

Ph.D. Thesis

by

Peter Fogh Odgaard

Department of Control Engineering
Aalborg University
Fredrik Bajers Vej 7C, DK-9220 Aalborg Ø, Denmark.

ISBN 87-90664-19-1
Doc. no. D-04-4750
September 2004

This thesis was typeset using LaTeX2$_\varepsilon$ in `report` document class.
MATLAB™is a registered trademark of The MathWorks, Inc.

Preface

This thesis is submitted in partial fulfilment of the requirements for the Ph.D degree at the Department of Control Engineering, Institute of Electronic Systems, Aalborg University, Denmark. The work has been carried out in the period from July 2001 to September 2004 under the supervision of Professor Jakob Stoustrup and Associate Professor Palle Andersen.
The Ph.D. project forms the part of the WAVES-project which deals with improvement of playability of Compact Disc players regarding surface faults based on wavelets or wavelet-like methods. This project has been conducted in cooperation with Bang & Olufsen A/S and Philips. The WAVES-project is supported by the Danish Technical Science Foundation (STVF) grant no. 56-00-0143.

Aalborg University, September 2004
Peter Fogh Odgaard

Acknowledgements

This thesis is the final product of the Ph.D project, and the work has been an exceptional experience. At times frustrating but fortunately the majority of the period has been a very positive experience. It has given me a possibility to do some interesting studies and research. In addition to this I have had pleasant experiences personally. I do especially think of my visits to St. Louis and Eindhoven. It has been three very good years, and I would like to thank all who made this work possible, helped and supported me during my Ph.D project.

First of all I would like to say thank you to my supervisors Professor Jakob Stoustrup and Associate Professor Palle Andersen, whom I have known both of you since I was an undergraduate Student and from this work I must say that you form a supervisor-team which is second-to-none. A former colleague who has been to great help is Enrique Vidal Sánchez, my predecessor as a Ph.D student working with CD-players, (now at Bang & Olufsen A/S). I would like to thank you for helping me with the experimental setup, and for the interesting talks and discussions concerning our projects. Also thank you to colleagues at the Dept. of Control Engineering and in the WAVES project, technical staff, and secretaries for support. Especially I would like to thank Karen Drescher for helping with transforming this thesis to be readable. I would also like to thank Henrik Fløe Mikkelsen from B&O A/S for practical help and for guiding the project in a relevant direction.

My stay abroad at Department of Mathematics, Washington University in St. Louis, USA, was a very pleasant experience both professionally and personally. I address my warmest thanks to my host Professor Mladen Victor Wickerhauser for help and guidance in my work, and from you Victor, Pia and I we learned what hospitality really is. I would also like to thank staff and faculty members at Dept. of Mathematics, Washington University in St. Louis for making Pia and me feel very welcome. Pia and I would also like to say thanks to an another person in St. Louis, who made our stay there very pleasant namely Irene Kalnins whom we got to know have learn to know through Rotary International. Thank you for taking care of us, and spending your weekends by showing us around in St. Louis.

I would like to thank the people I have met a couple of times at TU Eindhoven for some interesting discussions both professionally and privately when drinking a beer, es-

pecially Jan Van Helvoirt, Assistant Professor Pieter Nuij and Professor Maarten Steinbuch. I would also like to thank peoples I have met at Philips CFT and Philips Components Marcel Heertjes and George Leenknegt, and those at Philips Research who made our small but fruitful project possible: Henk Goossens and Koos Den Hollander. We have had some interesting discussions and cooperation also when finishing the project.
Also I my warmest thanks to my family Hanne, Ole, Rikke and Olga for their moral support somehow. I would also like to thank my brother-in-law Jesper for proof reading this thesis.
Finally, but certainly not least, dear Pia I would like to thank for infinite support during these three years. I know it has not always been easy to live with me and my project. It was wonderful that you stayed with me in St. Louis. A big moment during these three years, was on the CDC03 conference on Maui, looking on the sunset on the beach, where I proposed to you, and you replied : “JA” (YES).

Abstract

Several new types of optical disc standards have reached the market during the past years, from the well known Compact Disc (CD), and its computer version CD-ROM, through the Digital Versatile Disc (DVD) to the new high density discs, eg. Blu-Ray discs. All these standards have potential problems with surface faults like scratches and fingerprints. Many users have suddenly encountered the problem that a CD-player, a PC or a DVD-player could not play a given disc, since this disc has become too faulty. The main topic of this thesis is to improve the CD-players' capability to play these discs with surface faults, by improving the two controllers which keep the CD-player positioned on the CD. Some parts of this Ph.D project has been specialised for a specific CD-player. However, the achieved results can be generalised to the other optical disc player standards. The problem in handling these surface faults is that they result in a faulty signal component in the position signals which are used for positioning the CD-player. The basic idea behind this study is to use feature extraction to remove this faulty signal component from the measured position signals.

This Ph.D study contributes with six major technical contributions, which are all described in this thesis. Traditionally the optical parts of CD-players are modelled with simple linear models. However, these models are not ideal for detecting and handling surface faults. Instead a more detailed non-linear model of the optical part of the Compact Disc player is developed. Based on this new optical model and a simple model of the surface faults, a pair of residuals are defined. These residuals have shown better fault detection properties compared with the normally used residuals. In order to compute these new residuals a method to compute the inverse map of the non-linear optical model is introduced. Using this inverse map together with a Kalman estimator, the residuals can be estimated as well as an estimate of the position signals which are more valid during surface faults.

Based on the residuals, time-frequency based methods are designed for detecting the surface faults. The results, however, have shown that a better strategy is to improve standard thresholding algorithms. Time-frequency based methods are used to extract features from the surface faults. First of all the faults are classified among different fault classes, and even more importantly, the faults are approximated by the use of time-frequency based methods. These approximations are used for two purposes. The first

one is to simulate surface faults based on statistics of surface faults. The second usage of the approximation is in the feature based control strategy which removes the influences from the fault from the position signals by subtracting an estimate of the fault from the position signals. The feature based control strategy is tested by simulations and by experimental work. The results are conclusive in showing that a clear improvement is achieved, since the use of the feature based control scheme results in an improvement of the controller's ability to not to react on the tested scratches.

Resume

I løbet af de seneste år er adskillige nye optiske disk-standarder blevet lanceret, fra den bedst kendte Compact Disc (CD), og dens computer version CD-ROM, via Digital Versatile Disc (DVD) til de nyeste højtætheds diske, såsom Blu-Ray. Overfladefejl på diskene, bl.a. ridser og fingeraftryk, er et problem for alle disse typer af diske. Mange har netop oplevet dette problem med deres CD-afspiller, PC eller DVD-afspiller, at disse ikke har kunne afspille en disk fordi den havde overfladefejl. Hovedemnet i denne afhandling er hvorledes optiske diske med overfladefejl kan afspilles bedre. I dette arbejde er en bestemt CD-afspiller benyttet, men resultaterne forventes at kunne generaliseres til de andre optiske disk-afspillere. Problemet i at håndtere disse overfladefejl er, at overfladefejlene introducerer fejlsignaler på servosignaler. Den basale ide i dette projekt er at benytte feature extraction til at fjerne disse fejlsignal komponenter fra de målte servosignaler.

I denne Ph.D-afhandling er 6 tekniske bidrag præsenteret. Traditionelt er den optiske del af CD-afspillerne blevet modelleret med en simpel lineær model. Denne model er ikke ideel. I stedet for er en mere detaljeret model af den optiske del af CD-afspilleren udviklet. Et nyt par residualer er defineret baseret på denne optiske model og en simpel model af overfladefejlene. Residualerne har vist betydelige forbedringer mht. fejl-detektion sammenlignet med de normalt brugte residualer. For at beregne disse nye residualer er det nødvendigt at beregne den inverse afbildning af den ikke lineære optiske model. Benyttes den inverse afbildning sammen med en designet Kalman estimator kan residualerne estimeres.

Tids-frekvens baserede metoder er designet for at detektere overfladefejl i residualerne. Disse metoder resulterer i filtre, der uheldigvis bliver for snævre i frekvensbåndet til, at de kan benyttes til at detektere alle de overfladefejl metoden er testet på. I stedet har eksperimentale resultater vist, at en bedre strategi er at forbedre en almindelig tærskelværdimetode. Tids-frekvensbaserede metoder er derimod benyttet med succes i udtrækning af features fra de fejlbehæftede signaler. Først og fremmest benyttes de til at klassificere fejlene i ligeledes definerede fejlklasser, dernæst benyttes de til at approksimere fejlkomposanterne med. Disse approksimationer benyttes til to formål. Den første er at lave syntetiske ridser til brug i en simuleringsmodel, der simulerer regulatorernes evner til at håndtere CDer med overfladefejl. Det andet formål, er den “featurebaserede-

regulering strategi", der fjerner indflydelsen fra overfladefejlene i servosignalerne ved at subtrahere approksimationerne fra de målte servosignaler. Denne metode er verificeret ved simulationer og eksperimentelt arbejde, med det resultat at metoden medfører en klar forbedring af servoernes evne til ikke at reagere på de testede ridser.

Contents

Nomenclature

Acronyms

BRD	Blu-Ray Disc
CD	Compact Disc
CD-ROM	Compact Disc- Read Only Memory
CIRC	Cross Interleaved Reed-Solomon
DPD	Differential Phase Detection
DTD	Differential Time Detection
DVD	Digital Versitale Disc
EFM	Eighteen-to-Fourteen Modulation
FDI	Fault Detection and Isolation
HD	High Density
LQG	Linear Quadric Gaussian
OPU	Optical Pick-up Unit
PC	Personal Computer
PID	Proportional, Integral, Derivative controller
PLL	Phase Lock Loop
RC	Reed-Solomon
TOC	Table Of Contents
WAVES	Wavelets in Audio Visual Electronic Systems

WORM	Write Once Read Many

Constants/Variables

$\gamma_{\text{beg,low}}$	Low threshold for detecting the beginning of the surface faults, which is used if a fault is located.
γ_{beg}	Threshold for detecting the beginning of the surface faults.
γ_{end}	Threshold for detecting the end of the surface faults.
λ_{laser}	Wave length of the laser
$\mathbf{B}$	Flux density of the magnetic field.
μ_{ad}	Air/disc refractive index
ρ_{land}	Reflection rate of a land.
ρ_{pit}	Reflection rate of a pit.
A_{CD}	Area of the laser spot covering area around the information track.
a_{k}	Physical offset from main to side spots.
A_{l}	Area of the laser beam at the lens what passes through.
A_{track}	Area of the laser spot covering the information track.
b	Damper constant.
c	A constant.
d_{track}	Diameter of the focused laser spot on the information track
F_1	Focus point near the source and detectors.
k	Spring constant.
K_{pd}	Power driver gain.
L	Inductance in the coils
l_{u}	Distance from source to the lens.
l_{x}	Dobbelt distance from lens to the disc surface.
NA	Numerical Aperature
R	Resistance in the coils

Definitions

$\mathbf{f}(e_{\mathrm{f}}[n], e_{\mathrm{r}}[n])$	Vector function defining the optical model.
$\mathbf{f}_{\mathrm{e}}(e_{\mathrm{f}}[n], e_{\mathrm{r}}[n], \beta_{\mathrm{f}}[n], \beta_{\mathrm{r}}[n])$	Vector function defining the combined optical and fault model.
$\mathbf{g}$	Details wavelet filter.
$\mathbf{h}$	Approximation wavelet filter.
$\mathbf{I}$	Identity matrix.
$\mathcal{D}$	Disturbance set.
Φ	Continous scaling function.
$\Phi_{j,n}$	Discrete scaling function at scaling$j$ and translation n.
Ψ	Continous mother wavelet.
$\Psi_{j,n}$	Discrete mother wavelet at scaling$j$ and translation n.
$\tilde{\mathbf{g}}$	Dual details wavelet filter.
$\tilde{\mathbf{h}}$	Dual approximation wavelet filter.
$\mathbf{e}^{\mathrm{L}}$	The lifted distance vector.
$\mathcal{DP}$	Fault handling performance.
$\mathcal{K}(\cdot)$	Informational cost function.
$\mathcal{NP}$	Nominal performance.
$\mathcal{P}(\cdot)$	The feature based control scheme.
$\mathcal{P}^{\mathrm{L}}$	The lifted feature based control scheme.
Δ	One revolution delay.
ν	Lifting operator.
$AFC(n)$	Averaged frequency change function at sample n.
G_1	Class of small scratches.
G_2	Class of disturbance-like faults.
G_3	Class of large scratches.

$IFC(n)$	Instantaneous frequency change function at sample n.
K	Nominal controller.
P	Period of revolution.
S	The sensitivity of the nominal servo system.
S^{L}	The lifted sensitivity of the nominal servo system.
T	Complemtary sensitivity of the nominal servo system.
T^{L}	The lifted complemtary sensitivity of the nominal servo system.

Signals

$\alpha[n]$	Vector of the scaling residuals.
$\beta[n]$	Vector of the scaling fault parameters.
$\check{\mathbf{e}}[n]$	Vector of surface fault components in the distance signals.
$\ddot{\mathbf{x}}(t)$	Acceleration of the OPU.
$\dot{\mathbf{x}}(t)$	Velocity of the OPU.
$\grave{\mathbf{e}}[n]$	Vector of the cleaned distance signals.
$\hat{f}[n]$	Dynamic estimate of the discrete time signal $f[n]$
$\mathbf{d}_{\mathrm{m}}[n]$	Vector of the mechanical disturbances.
$\mathbf{d}_{\mathrm{s}}[n]$	Vector of the self pollutions.
$\mathbf{e}[n]$	Vector of distances from the OPU to the track.
$\mathbf{f}[n] = \begin{bmatrix} \mathbf{f}_{\mathrm{f}}[n] \\ \mathbf{f}_{\mathrm{r}}[n] \end{bmatrix}$	is a vector of the focus and radial signal components.
$\mathbf{f}_{\mathrm{s}}[n]$	Vector of surface faults.
$\mathbf{r}[n]$	Vector of the orthogonal fault parameters.
$\mathbf{s}[n] = \begin{bmatrix} D_1[n] & D_2[n] & S_1[n] & S_2[n] \end{bmatrix}^T$	Vector of the four photo detector signals.
$\mathbf{s}_{\mathrm{m}}[n]$	Vector of measured detector signals.
$\mathbf{u}[n]$	Vector of control to the OPU.
$\mathbf{x}[n]$	Vector of the position of the OPU.

$\mathbf{e}_1$	Vector of the simulation output in the surface fault free case.
$\mathbf{e}_2$	Vector of the simulation output in the case of the surface faults.
$\mathbf{k}_\mathrm{f}$	Focus correction block.
$\mathbf{k}_\mathrm{r}$	Radial correction block.
$\mathbf{x}(t)$	Position of the OPU.
$\tilde{f}[n]$	Static estimate of the discrete time signal $f[n]$
$f(t)$	Continous time signal
$f[n]$	Discrete time signal
$\mathbf{a}_j[n]$	Wavlet approximation at scale j.
$\mathbf{d}_j[n]$	Wavlet details at scale j.
$\mathbf{f}_\mathrm{d}[n]$	Vector of fault detection signals.

List of Figures

List of Tables

Chapter 1

Introduction

Compact Disc (CD) players have been on the market since the beginning of the eighties, and now every home has at least one CD-player. The development of CD-players have been carried out through a number of different research areas. One of these areas is control engineering. This research area together with signal processing is the main part of this thesis, and these areas are used to improve the CD-player's aviability to play CD with surface faults like scratches.

In Section 1.1 the problem of handling surface faults in CD-players are shortly described and motivated. A overview of previous related work is given in Section 1.2. The problem dealt with in this project and how it is done is described in Section 1.3, the contributions of this Ph.D project are presented in Section 1.4. An outline of this thesis is given in Section 1.5.

1.1 Motivation of the problem

The music is retrieved from the CD by the use of an Optical Pick-Up. This Optical Pick-Up has no physical contact with the data track on the CD. Therefore control is needed for the Pick-Up to follow the track and to focus the laser beam at the track. By use of smart optics the Pick-Up generates approximations of position errors of the Pick-Up relative to the track. This sensor strategy is very usable in cases of healthy CDs, but if surface faults occur on the disc, these faults will influence the position sensor signals. In other words during surface faults it is not a good strategy to rely on the position sensors, or at least fully rely on them. This is in conflict with the specification of the controller for handling disturbances like mechanical shocks etc, which the controller should suppress meaning that a high bandwidth is needed. The idea of this Ph.D project is to solve this non-trivial problem of handling both surface faults and disturbances at the same time.

1.2 Overview of previous and related work

CD-players have been on the market since 1982, where Philips and Sony launched their first players. However the publication of research results on control of CD-players was delayed with a decade. From the beginning of the nineties research in control of CD-players has been intense, especially in the usage of adaptive and robust controller. [Steinbuch et al., 1992] is the first example on a μ-controller to a CD-player based on DK-iterations. An example of an adaptive control design is [Draijer et al., 1992] where a self-tuning controller is suggested. In the following years a large number of different control strategies were applied to the CD-players. [Dötch et al., 1995] suggested an adaptive repetitive method, quantitative feedback theory is suggested in [Hearns and Grimble, 1999], rejection of non/repeatable disturbances is suggested in [Li and Tsao, 1999], fuzzy control is used in [Yen et al., 1992], hybrid fuzzy control is used in [Yao et al., 2001], linear quadratic Gaussian control is used in [Weerasooriya and Phan, 1995] and disturbance observer is used in [Fujiyama et al., 1998]. [Wook Heo and Chung, 2002] uses vibration absorber to damp the mechanical disturbances.
The development of DVD-players has implied some attention to this application. In [Zhou et al., 2002] sliding mode control is used to improve the performance against mechanical disturbances if compared with a traditional PID controller. [Filardi et al., 2003] applied a robust control strategy to the DVD-player. [Zhu et al., 1997] uses iterative learning control to perform the radial tracking in a DVD-player.
The research is still focused on the CD-player application, this is illustrated in Ph.D dissertations finished from the late nineties to now. [Dötsch, 1998] studied system identification for control design in CD-players, and in the same year [Lee, 1998] used CD-players as a case study in his study of robust repetitive control. More recently, [Dettori, 2001] used LMI techniques to solve: multi-objective design and gain scheduling. [Vidal Sánchez, 2003] dealt with two problems, uncertainties of the CD-player by the use of μ-control and implementation of a controller tolerant towards surface faults. A clear majority of the research done in the area of optical drives is focused on minimising the disturbances channel from mechanical disturbances/disc deviations to the position error. Only [Vidal Sánchez, 2003] addresses the problem of handling surface faults such as scratches, fingerprints etc. The surface faults impose an upper limit on the controller bandwidth, which is in conflict with the minimisation of the disturbances channels, since minimisation of the disturbances require high controller bandwidths. [Heertjes and Sperling, 2003] and [Heertjes and Steinbuch, 2004] indirectly handle the surface faults by the use of non-linear filters to improve controller sensibility without making the controllers more sensitive towards the surface faults.

1.2.1 Background at Aalborg University

In 1997 Department of Control Engineering at Aalborg University and Bang & Olufsen started a co-operation in advanced control of CD-players. The start was a master stu-

dent project, see [Vidal et al., 1998]. This student project showed the need for a larger research project (OPTOCTRL). OPTOCTRL dealt with the use of advanced control techniques for improving the reproduction of sound in CD-players. This was the subject of Enrique Vidal Sánchez's Ph.D thesis [Vidal Sánchez, 2003], and in a number of student projects. OPTOCTRL did also host a project of low cost optical active sensors, see [la Cour-Harbo, 2002]. In this work joint time and frequency based methods were used to improve optical detectors. In [Vidal Sánchez, 2003] it was concluded that the use of advanced control techniques can improve quality of reproduced sound in case of faults on the CD surface.
In 2001 another research project was founded partly based on OPTOCTRL. This project called Wavelets in Audio/Visual Electronics Systems (WAVES), contains a work package handling surface faults on CDs with the use of joint time/frequency based feature extraction to improve the advanced controllers. This Ph.D project that makes the basis of this thesis is the work packages (Feature based control of Optical Disc players).

1.3 Basic idea of the Ph.D project

The idea is to handle the surface faults such as scratches and fingerprints without decreasing the suppression of disturbances. This can be achieved by viewing this control problem of handling surface faults in CD-players as a fault tolerant control problem, where the surface faults are viewed as fault which can be handled by the use of an active fault tolerant controller, where the surface defects components are removed from the measurement signals. This adaption are done by the use of feature extraction of the defect, where time-frequency analysis method are used to extract these features, see [Mallat, 1999].

1.4 Contributions

In the following list the main contributions of the author are summarised, most of these contributions have been or are going to be published in international conference proceedings and international journals.

- A model of the optical detector system, including cross-couplings between focus and radial detectors see [Odgaard et al., 2003c]. This model is non-linear and shows a clear coupling between focus and radial detectors.
- A method for computing the locally defined inverse map of the optical model is presented in [Odgaard et al., 2004a]. This computes a static estimates of focus and radial positions and decoupled residuals for detection of surface faults.
- A Kalman estimator is used to compute dynamical estimates of focus and radial positions based on the static estimates, see [Odgaard et al., 2003b] and [Odgaard et al., 2003a].

- An advanced version of the threshold method is developed in order to improve the time localisation of the surface faults, [Odgaard and Wickerhauser, 2004], based on the decoupled residuals. In addition two time-frequency based methods are designed based on Fang's algorithm, [Odgaard and Wickerhauser, 2004], and wavelet packet filters, [Odgaard et al., 2004e].

- A local discriminating basis is found, the most discriminating basis vectors of this basis are used to discriminate among different classes of faults, [Odgaard and Wickerhauser, 2003]. The members of these defect classes can be approximated by use of a few most approximating Karhunen-Loève basis vectors [Odgaard and Wickerhauser, 2003]. The coefficients as well of these basis vectors are the extracted features.

- A model of surface faults is formed see [Odgaard et al., 2003a] and [Odgaard et al., 2004d]. This model can be used to design Fault Tolerant Controllers, and to simulate CD-player playing a disc with surface faults [Odgaard et al., 2004d]. This model is implemented in MATLAB™, [Odgaard et al., 2003d].

- A Fault Tolerant Control scheme is developed. This scheme handles surface faults by the use of the extracted features [Odgaard et al., 2004c] and [Odgaard et al., 2004b].

1.5 Outline

The remaining parts of the thesis are organised as follow:

Description of Compact Disc players introduces the Compact Disc technology for the reader. Other optical disc formats are also introduced to give an idea of why the result achieved at the CD-players can be generalised to other disc formats. The Optical Pick-Up and positioning servos are described and explained. Finally the interesting control problems in the CD-player are discussed and explained. This leads to a proposed structure of the feature based controller scheme.

Model of Compact Disc players introduces and describes the Test Setup used in this work. This is followed by a description of the overall model structure and model of CD-player containing the dynamical model of the electro-magnetic part, the non-linear static model of the optical part and a model of surface faults.

Fault Residuals based on Estimations of the Optical model derives a local solution to inverse map of the combined optical model of the CD-player and model of the surface faults. The solution gives estimates of the fault parameters (residuals) and the actual position of the optical pick-up.

Time-Frequency analysis introduces some relevant time-frequency theories and bases, these theories and bases are of large importance in the remainder of this thesis.

Time localisation of surface faults describes three developed methods for locating and detecting the surface faults. These methods uses the new derived fault residuals as the basis of the localisation and detection of the surface faults. The three methods are based: on Fang's algorithm for segmentation of the time axis, joint best basis wavelet packet filters and en extended version of the normally used thresholding method. These methods are all validated by the use of experimental data.

Time-frequency based feature extraction of Surface Faults describes the general idea of extracting features, especially is given an explanation of which features are interesting for the control of the CD-player. Two different kinds of features are extracted: class of fault, and approximating coefficients.

A Simulation model develops a simulation model of CD-players playing discs with defective surfaces. This simulation model is based on a model of a CD-player, the model of surface faults, and time-frequency based features extracted in the previous chapter.

Feature based control of Compact Disc players describes the basic idea of this feature based control scheme. The scheme is related to the general fault tolerant control scheme. Moreover this control scheme is derived based on the extracted features of surface fault. These features are used in order to remove the faults influence on the positioning servos. Stability issues of the scheme is discussed and the scheme is proven to be stable. Finally the scheme is validated through simulations and experimental tests.

Conclusion gives concluding remarks on this thesis and some suggestions for future work in this field.

Chapter 2

Description of Compact Disc players

In this chapter the optical disc player in general and the CD-player in particular are described. This description makes it possible to identify the challenging control problems in the optical disc players. This leads to a description of the main topic in this Ph.D project: handling of surface defects based on a feature based fault tolerant control scheme. In Section 2.1 the different optical disc formats are described, this leads to a description of the optical discs, see Section 2.2. In Section 2.3 the optical disc player is described. Some basic optical principles are the topics in Section 2.4. These principles are used to compare the different kinds of optical disc formats, which are followed by the basics in the optical pick-up, see Section 2.5. In Section 2.6 some of the most used optical measurement principles are described. The servo loops are described in Section 2.7. In Section 2.8 the performance requirements, disturbances and defects are discussed. Finally, in Section 2.9 the specific fault tolerant control scheme used in this work is described.

2.1 Optical disc formats

Even though this thesis is focused on the Compact Disc (CD), the most well known optical disc formats will shortly be described and compared to the CD. The first optical data storage technique was announced by Philips in 1972. The first optical discs were analog optical discs see [Bouwhuis et al., 1985]. Almost a decade later in 1981 Philips and Sony proposed a digital standard, the Compact Disc Digital Audio standard (CD-DA), which was coded in the so-called red book [Philips and Sony Corporation, 1991]. The year after in 1982 Philips and Sony launched their first CD-players on the market. Almost 10 years later in 1991 the sale of CDs exceeded the sale of audio-cassettes, and

the long play records, its predecessor as the high quality music storage media.
In addition to audio application of CDs, the CDs have also been developed in other versions for other applications. The most well known other application is for binary data storage on computers, CD-ROM. Another application is CD-video where the CD media is used to storage videos, movies etc. The CD-videos have never conquered a large part of the video market, since one disc with the in the standard used compression cannot contain an entire movie. To day with the use of standard computers and new compression algorithms it is possible to store an entire movie on a CD-ROM. Todays fast computers can decompress and play the stored movie in real-time.
The next generation of optical disc players, the Digital Video/Versatile Disc (DVD) has the capacity to store an entire movie in a good quality on one disc. A CD has the storage capacity of 650/700 MB and the DVD upto 25 times as much. The DVD can be used for the three applications audio, video and ROM as well as the CD.
The generation of optical discs which are the successors of the DVD is now in the phase of development, this type of optical discs are typically called High Density Optical Discs. For the time being four different standards have been proposed, with a typically storage capacity from 20 GB to 50 GB. (Approximately 30 to 75 times the CD capacity). These four standards are: the Blu-ray disc (BRD)[Hitachi, Ltd. et al., 2002], the advanced optical disc [Toshiba Corporation and NEC Corporation, 2002], the Blue-HD (High Density) disc [AOSRA, 2003] and the HD-DVD-9 disc [Warner Bros, 2002]. They are all re-writable discs.
In addition to the large variety of disc types, the CD and DVD are available in a mass replicated moulded disc, Write Once Readable Memory (WORM), and re-writable discs. These different types of discs are shortly described below.

Mass replication moulded discs These discs are prerecorded CDs or DVDs containing audio, video or data. The information is stored on these discs in a mass duplication process by the use of stamping. This stamped surface is following coated with an ultra thin layer of reflective material such as copper, aluminium, gold or silver. The production of the disc is finalised with a layer of transparent protection material. It is clear from the description that this disc cannot be re-recorded.

WORM discs These Write Once Read Many times discs are a group of discs containing a recordable CD (CD-R) and a recordable DVD (DVD-R). These discs are in many ways much alike the first group of discs, but deviate in one important way. They have a layer of photosensitive dye covered by a reflective metallic layer. The discs are recorded by burning the photosensitive material in the dye by a laser. This process is not reversible and only durable once, meaning this burning is permanent.

Re-writable discs In these discs the WORM discs' photosensitive dye layer is replaced with a special phase-change compound. The phase of this compound can be

changed due to the amount of energy applied by the laser. This means that it is possible to erase and rewrite the disc. These discs lifetime measured in numbers of rewritings varies from a 1000 times (CD-RW, DVD-RW, DVD+RW) to 100,000 times (DVD-RAM). All the proposed High Density discs are in principle re-writable. However, ROM versions of the High Density discs are being discussed.

These descriptions give a brief overview of the wide variety of optical discs on the market, this overview is illustrated in Fig. 2.1. The competition of the leading position on the re-writable DVD market has not declared a winner yet. One could say that the competition on the High Density disc marked has not really started, since the competing standards have not been completely established. The optical principles behind all these discs are similar. In the following sections these principles are described by focusing on CDs. During this description some of the important differences between the different kinds of discs are mentioned.

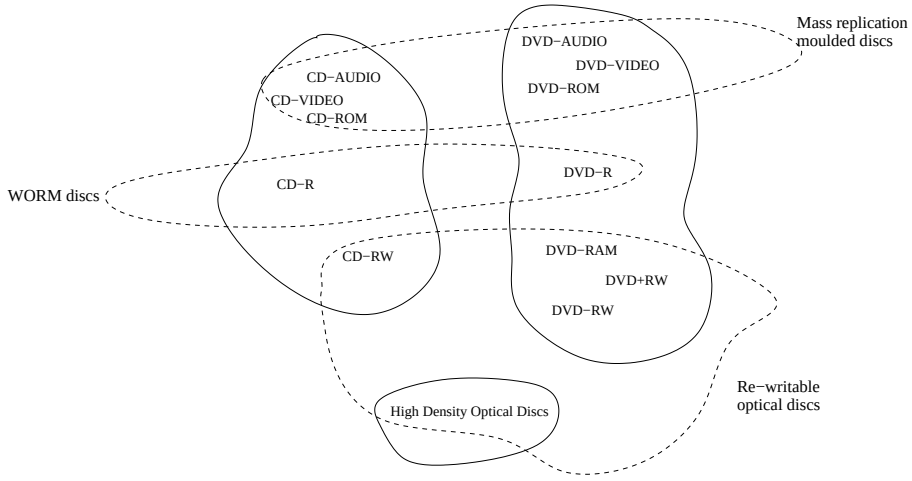


Figure 2.1: The most common types of optical discs. CDs, DVDs and HDs grouped into disc production types.

2.2 The optical disc

The diameters of the CD and the DVD are for both 120 mm, some of the HD discs are smaller, and some of them do have a diameter on 120 mm. The data is on all discs stored in a clockwise spiral track starting from the centre of the disc. The spiral consists of microscopic pits. The length of the pits and the area between them form the recorded

information, the area in track between pits is called lands. The information in the track consists of three parts. The first part starting from the centre of the disc is the lead-in area, which contains a table of contents of the disc (TOC). The area following the TOC is the program area containing the stored information. The track is finished by the end of the program area which is the lead-out area of the disc. The general structure of the disc is illustrated in Fig. 2.2.

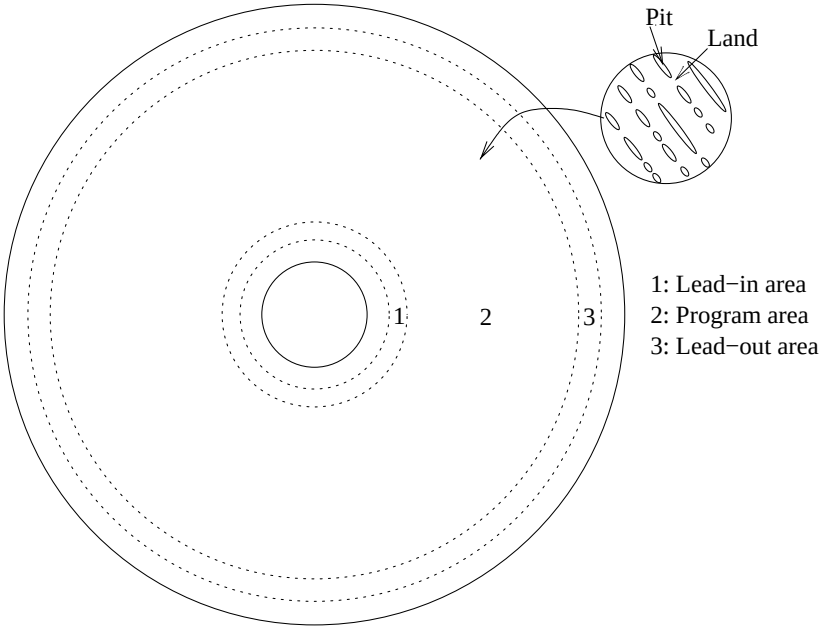


Figure 2.2: General structure of optical discs illustrates the placement of the three areas: lead-in, program and lead out.

The basic material in the manufacturing of optical disc is polycarbonate plastic. This inexpensive material protects and hosts the information layer and its reflective coating. A cross section of an optical disc is illustrated in Fig. 2.3. Seen from the readout side the laser will first meet the optical transmittance protection layer, through which the laser will be focused. In addition to the focusing of the laser this layer has a purpose related to its name: It protects the information layers from being damaged by sharp objects. However, the protection layer will often be damaged instead, which causes non intended focusing and transparency properties in the damaged area, (this problem will be described in more details later in this chapter). The information layer is coated with a reflective material, since it has poor reflection properties itself. The reflected laser will pass back through the protection layer and be detected in the optical detectors

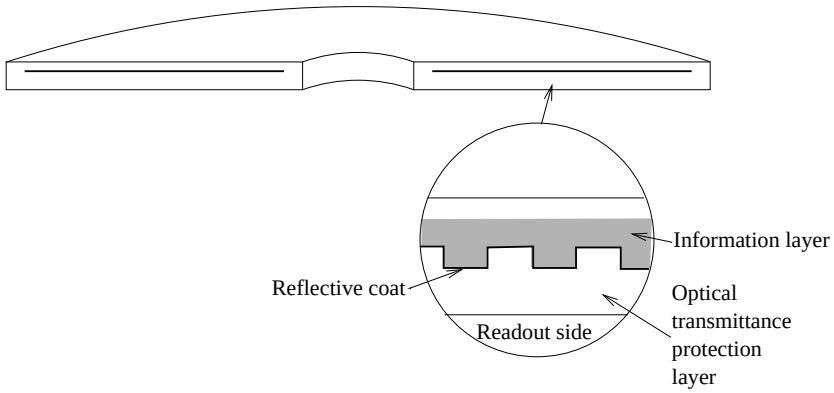


Figure 2.3: Cross section of the disc. This disc has only one information layer as CDs, even though some other optical discs have several information layers.

(optical sensors). In dual layer DVDs the upper information layer is covered with a semi-reflective material, through which it is possible to focus the laser. In addition the upper layer has a lower density compared to a conventional layer such as the lower layer. By using this lower density the Signal Noise Ratio (SNR) can be kept low. The side opposite the readout side is normally called the label side where a label can be printed on the disc. But in case of a double sided discs this label has to be transparent for the laser beam.
Some interesting key specifications of the different generations of optical discs are compared in Table 2.1, based on [Philips and Sony Corporation, 1991], [ECMA, 2001], [Hitachi, Ltd. et al., 2002]. As a representative of the group of High Density optical discs the Blu-Ray Disc (BRD) is chosen. Other standards in this group may have other specifications. All the HD discs standards except HD-DVD-9, increase the information density by the use of blue lasers. The HD-DVD-9 increases the density by use of a more efficient data coding algorithm.

2.3 The optical disc player

A block diagram of the optical disc player is shown in Fig. 2.4. This block diagram illustrates the most important parts of an optical disc player. The OPU is the Optical Pick-up Unit which emits and focus the laser at the disc surface, it also detects the laser beam reflected back from the information layer in its photo detectors. The remaining system can naturally be divided into three subsystems: the data path, control/servo system and the logic block. The logic block serves the role as being interface between the user and the servo part and performs the required logic in order to keep the correct sequences of operation. The purposes of the two other subsystems can shortly be described. The

	CD	DVD	BRD
Disc Diameter [mm]	120	120	120
Disc thickness [mm]	1.2	1.2	1.2
Tracking pitch [μm]	1.6	0.74	0.32
Optical transmittance protection layer [mm]	1.17	0.6	0.1
Laser wavelength [nm]	780 (infrared)	650 (red)	405 (blue-violet)
Numerical aperature (NA)	0.45	0.6	0.85
Air/disc refractive index (μ_{ad})	1.55	1.55	1.55
Data layers	1	1 or 2	1 or 2
Readout sides	1	1 or 2	1 or 2
Data capacity [GBytes]	0.65 or 0.7	4.7-17.0	23.3-50

Table 2.1: The most interesting specifications in the CD, DVD and BRD standards. The BRD has been chosen as an example of the High Density Disc standards, as the specifications are available at this time.

control/servo system has to positing the OPU such that it can generate some signals depending on the information stored in the information layer. The job of the data path is to convert these retrieved signals into the data stored in the information layer. The sledge servo moves the OPU for coarse radial adjustments and focus and radial servos are used for fine positing of the OPU. The disc motor has the important function to spin the disc around at the adequate speed.

The OPU generates the self-clocking waveform data from the disc. This signal is referred to as High Frequency (HF) signal. This signal can be coded in different ways depending on the type of disc. In CDs the Eight-to-Fourteen Modulation (EFM) channel code is used for coding the data. 8 bit data are represented on the disc by 14 channel bits plus 3 additional bits. The length of the pits and lands is in the interval from 3 channel bits to 11 channels bits, see [Stan, 1998]. The DVDs use an advanced version of the EFM coding called EFM+, where the 8 data bits are represented by 17 channel bits on the disc. The Data Separator separates the HF signal into: subcode bytes, data samples, clock signal. The subcode is a kind of a time stamp on each data sample, which is used by the logic unit to locate the required data samples in the correct order. The PLL lock is used to decode the EFM signal, for more information see [Stan, 1998].

The separated data samples are fed to the Error Corrector, in which the redundant data are used to check and eventually correct erroneous data. The error correcting method in CDs is Cross Interleaved Reed-Solomon code (CIRC). This error correction method is based on the work by Reed and Solomon, see [Reed and Solomon, 1960]. Its maximum correction length is 4000 bits which are approximately 2.5[mm] track. The Reed-Solomon (RS) product code, a variant of the CIRC, is used for error correction in DVDs, is capable of handling errors relating to approximately 6[mm] track. The deinterleaver

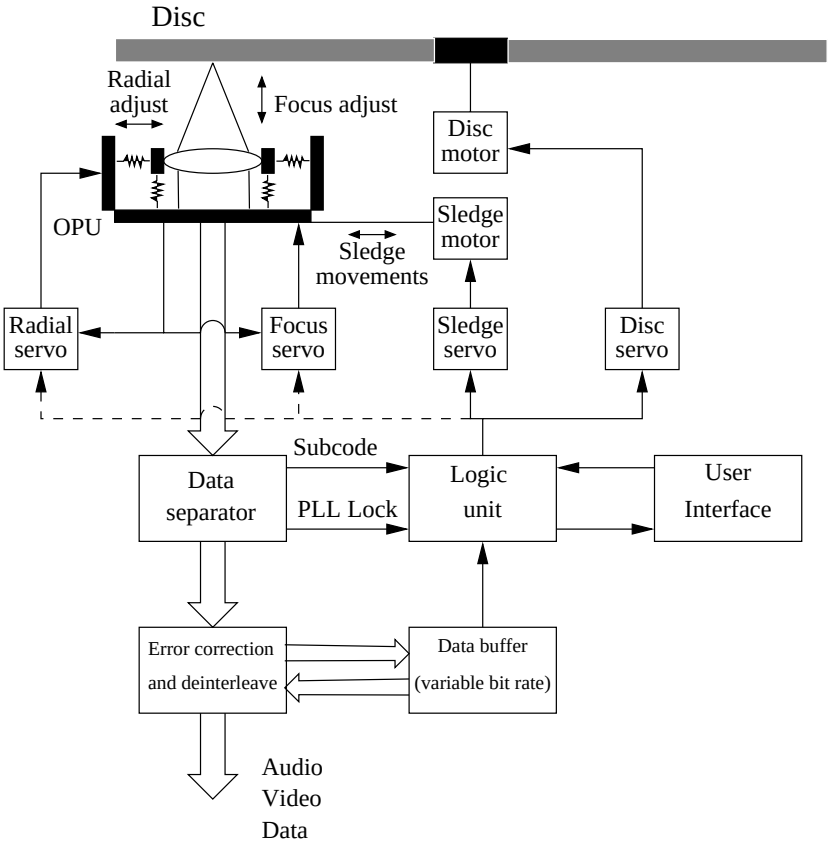


Figure 2.4: A block diagram of general structure of the optical disc player. The broad arrows illustrate the data path, and the narrow arrows handling and control. The dashed arrows are logical signals which handles special operations, such as start-up, track jump etc. The OPU is the Optical Pick-up Unit which is used to emit and detect the laser beam.

writes samples sequentially into the data buffer and read it out again by the means of a sequencer. The rotational speed of the disc motor is controlled by the size of this data buffer. If the buffer is almost full the disc motor decreases its speed and if the buffer is almost empty the speed of the disc motor is increased.

2.4 Optical principles

It is difficult to study the relief structure formed by the microscopic pits in the information layer by use of conventional optics. In 1934 Zernike described a technique called contrast microscopy, see [Bouwhuis et al., 1985]. This technique is the basis of optical interferences techniques which today are used to read these relief structures.
A laser beam (**l**ight **a**mplification by **s**timulated **e**mission of **r**adiation) is light composed by photons bouncing synchronously. This means that these photons have the same well defined frequency and wavelength λ_{laser}. In optical disc players the laser is focused by lenses and through the optical transmittance protection layer and further to the information layer.
The core idea in the optical interference techniques is the fact that the difference in level between the pits and lands is a quarter of the wavelength of the laser inside the protection layer. The area surrounding the track has the same level as the land. The radius of the laser spot focused at the information track has such a size, that the spot covers an area outside the track. In the case it covers land and surrounding area, all reflected light will be reflected with the same phase. In the other case, where the spot covers a pit and some surrounding areas, the light reflected from the pit has 180 degrees extra phase shift, meaning that light reflected from the pit will annihilate the light reflected from the surrounding area. This means that amount of light leaving the disc is dramatically reduce in this situation. The reflected light is detected by a number of photo detectors placed on the path of the reflected light.
It is easy to verify that the difference in levels between the pits and lands is a quarter of wavelength of the laser. From Table 2.1, it is known that the wavelength of the laser is 780[nm], the change of media from air to the denser polycarbonate plastic reduced propagation speed of the light. The ratio of velocity (refractive index of the media), is $\mu_{\text{ad}} = 1.55$, this gives a reduction of the laser wavelength in the protection layer from 780[nm] to approx. 500[nm], which is approximately four times the pit height 120[nm].
The principles behind the phase contrast microscopy applied to optical discs are now described, with the purpose of investigate how is the information density increased on the disc. An indication to this answer can be given by estimating the effects of surface contamination. Some definitions and calculations are needed in order to answer the question. The angle, θ_{in}, of the incident beam entering the disc's protection layer

$$\theta_{\text{in}} = \sin^{-1}(NA), \tag{2.1}$$

where NA is the numerical aperture of the lens in the OPU. The angle of the refracted

	CD	DVD	BRD
Incident angle θ_{in} [°]	27	37	58
Refracted angle θ_{out} [°]	17	23	33
Readout beam size D_{beam} [mm]	0.72	0.51	0.13
Focused beam size at information layer d_{track} [μm]	0.87	0.54	0.24

Table 2.2: The calculated optical parameters for CD, DVD and BRD.

beam can be computed by

$$\theta_{\text{out}} = \sin^{-1}\left(\frac{\sin(\theta_{\text{in}})}{\mu_{\text{ad}}}\right), \tag{2.2}$$

where μ_{ad} is the refraction index which can be seen in Table 2.1. [Bouwhuis et al., 1985] gives the diameter of the focused laser spot at the information track, d_{track},

$$d_{\text{track}} = \frac{\lambda_{\text{laser}}}{2 \cdot NA}. \tag{2.3}$$

After computing the spot diameter at the information track it is possible by the use of geometric computations to compute the diameter of the readout beam, D_{beam}, at the disc surface. It is given by the following equation

$$D_{\text{beam}} = \frac{2 \cdot T_{\text{l}}}{\tan(90^\circ - \theta_{\text{out}})} + d_{\text{track}}\ [m], \tag{2.4}$$

where T_{l} is the thickness of the protection layer, and is given in Table 2.1.
The density of information on an optical disc can be approximated, see [Bouwhuis et al., 1985] by

$$\text{density} = \left(\frac{NA}{\lambda_{\text{laser}}}\right)^2, \tag{2.5}$$

(2.5) indicates two possibilities to increase the information density on the optical disc, either by increasing the NA and/or decreasing the wavelength of the laser beam. Increasing the NA and/or decreasing λ_{laser} results in a decreased focused spot size. These optical changes also introduce a side effect of an increased optical aberration (distortion), which is compensated by reducing in thickness of the optical transmittance protection layer. One consequence of the reduction of thickness of the optical transmittance protection layer, is clear. The maximum depth of a scratch in the disc surface without damaging the information layer is decreased as the thickness of the optical transmittance protection layer decreases, since the laser spot size at the disc surface also decreases surface defects would appear larger as well. To study this phenomenon more detailed the figures of the above principle calculations are performed for: CD, DVD and BRD, as representative of the high density optical disc, see Table 2.2.

Examples on some surface defects as dust particles and a human hair on the disc surface are by rule of thumb considered to be of the respective sizes: 40[μm] and 75[μm]. By comparing these sizes with D_{beam} of the different medias in Table 2.2 it is seen that a hair or a dust particle are not an obstacle in the readout on a CD, due to the high ratio between D_{beam} and the size of the defect. This ratio decreases from CDs to DVDs where these kind of defects tend to be an obstacle in the readout process, and for the BRD these defects are almost as large as the laser beam itself, and are thereby even a larger obstacle.

2.5 The optical pick-up

The standards of optical disc players do not describe much concerning the optical pick-up, they only give a set of requirements to the wave length of the laser beam and the numerical apperature of the lens. These few requirements to the optical pick-up give a large degree of freedom for the designers of the optical pick-up; as a consequence the general principles will be described, and are illustrated in Fig. 2.5.
The laser beam is emitted by the laser diode. The light beam will following meet an objective lens before it meets a polarising prism with a defined transmittion plane. The light is following phase shifted with a phase of 90° in the Quarter-wave plane. The laser beam is focus at the information layer by the use of the 2-axis moving objective lens; the light beam reflected back from the informations layer also passes through this lens. The quarter-wave plane again phase shifts the reflected light with 90°. The light beam is following reflected in the polarising prism towards the photo diodes (photo detectors). An objective lens is used to focus the light beam at the photo diodes. An alternative implementation is to implement all the optical elements, except the 2-axis moving lens, laser and photo detectors in a hologram. In addition the laser and the photo detectors can be placed in the same housing as the hologram, this implementation is often used since it is more environmentally and mechanically stable. The optical pick-up is in addition to readout of data also used to measure the focus and radial distances. The implementation of these indirect measurements variates from players and design; some of the most used principles are described in the next section. In the following the Optical Pick-up Unit, will be denoted as OPU.

2.6 Optical measurements

As a consequence of the lack of physical contact between the OPU and the optical disc surface, it is needed to measure the position of the OPU relative to the information track in two directions: in the focus direction and in the radial direction. There is a large variety in the methods and principles used for measuring these distances. In the following some of the most used ones are shortly described, starting with the focus distance measurements.

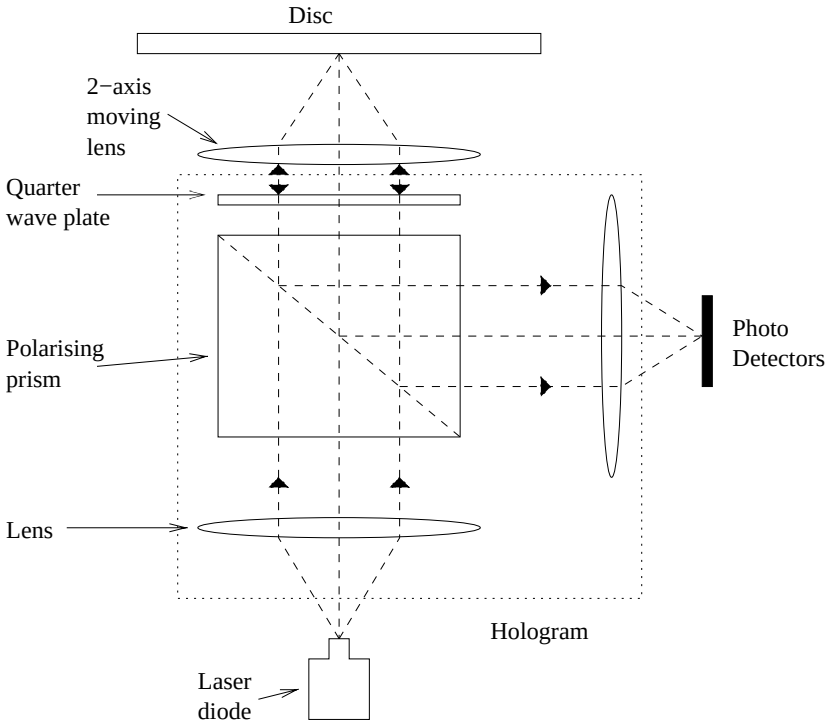


Figure 2.5: The general principle of the optical pick-up.

2.6.1 Focus distance measurements

The four most used focus distance measurements are shortly described, starting with the astigmatic principle, followed by single Foucault, double Foucault and spot-size detections. The general principles behind all these methods are to place some kind of asymmetry in the path of light reflected from the disc.

Astigmatic principle The astigmatic principle is illustrated in Fig. 2.6. A cylindrical lens is placed in front of the photo detectors. This lens has two focus points, one in front of the photo detectors and one behind them. The image of the laser beam on the photo detectors will be an ellipse whose aspect ratio changes as a function of the focus distances. In the cases of the focus distance being equal to zero, the image will be circular, since the cylindric lens is designed in such a way that zero focus distance is in between of the two foci of the cylindric lens. The focus distance is measured by dividing the photo detectors into four quadrants. When these are connected as shown in Fig. 2.6 the focus distance is computed. The data readout is the sum of the high frequency signal

from all these four photo detectors. The mapping from the focus distance to astigmatic measurement of the focus distances is shown in Fig. 2.8 where it is compared with the Single Foucault mapping.

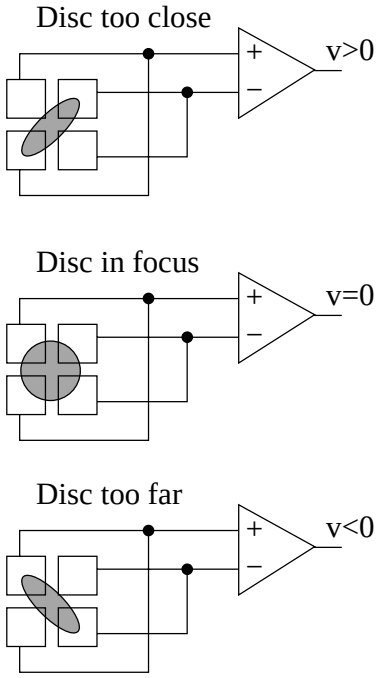


Figure 2.6: Illustration of the astigmatic principles. In the top an illustration where the disc is too close, in middle the disc is in focus and at the bottom the disc is too far away.

Single Foucault principle The Single Foucault principle is also called the knife-edge method since a knife edge introduces the asymmetry in the light path. The light beam will be detected by two photo detectors. The knife edge is placed in a position that gives both photo detectors the same level of light energy in the case of zero focus distance, as illustrated in Fig. 2.7. In the cases where the light beams is out of focus, the knife edge will change the ratio of detected light at the photo detectors, which is illustrated in Fig. 2.8. The absorption of light increases as the numerical focus distance increases.
Based on the description it is clear that the placement of the knife edge is highly important. Just a small misplacement can result in large measurement errors. This problem is often solved by implementing the knife edge in a hologram together with the lenses.
The mappings of the Single Foucault principle and the astigmatic principle are shown in Fig. 2.8. In principle these curves are different as in Fig. 2.8, see [Stan, 1998] and

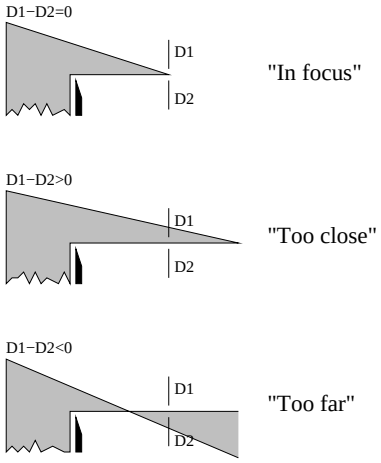


Figure 2.7: Illustration of the Single Foucault focus detector principle. For simplicity this illustration is based on point source laser, the principle in real world lasers is the same.

[Vidal Sánchez, 2003]. These assume, however, the Single Foucault photo detectors to be of infinite size. This is of course not the case in real applications. The designers will often choose to minimise these detectors. As the focus distance increases, the size of the laser spot both on the disc surface and photo detectors increases too. A consequence of this is that the amount of energy detected decreases dramatically towards zero, see [Odgaard et al., 2003c] and Section 3.4.

Double Foucault principle In this principle the knife edge is replaced with a prism. The prism is used to split the light beam along the optical axis. Two pairs of photo detectors are used to obtain the distance error signal based on the split laser beam. The Double Foucault principle is rarely used, it is replaced by its counterpart the Single Foucault principle, [Stan, 1998].

Spot-size detection principle This method also uses a prism and two split detectors, and is still used in some CD-ROM drives, [Stan, 1998]. This principle measures the change in spot size since the spot size depends linearly on the focus distance.

2.6.2 Radial distance measurements

The radial distance can be measured in a number of ways. In the following some of the most commonly used ones are shortly presented.

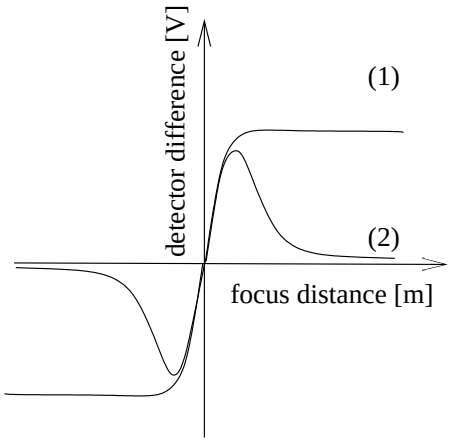


Figure 2.8: An illustration of the theoretical Single Foucault (1) and practically Single Foucault and Astigmatic, (2), optical mappings. e_{f} is the focus distance, u is the photo detector difference. Notice that these two method generates in theory different optical mappings, the practically Single Foucault mappings is alike the Astigmatic mapping.

Three beams principle In addition to the main beam used for data readout and focus distance measurements two additional laser beams are used. This method is also called twin-spot radial detection. The two additional laser beams are derived from the main laser beam by a grating structure, the side spots are placed aside the main beam, with an offset a_{k}, as illustrated in Fig. 2.9. If a spot is fixed over the track, the received high frequently photo detector signal will have an amplitude depending on the pits and lands in the track. However, a low frequently signal will be of constant value for a constant radial distance, since the sensor signals are low-passed filtered the pits and lands are not of any interest in this process of computing the radial distance. These can be removed from the photo detector signals by low-pass filtering these detector signals. When low-pass filtered, the radial error can be approximated by the difference between the two side photo detectors as illustrated in Fig. 2.9.

Track Wobble principle The idea behind this principle is to feed a sinusoidal signal to the radial servo, causing a radial oscillation of the radial distance of a few decades of nanometers. This modulates the envelope of the readout signal. This modulation can be used to obtain the radial distance.

Push-pull principle The push-pull principle uses a diffraction pattern to analyse the reflected beam. In the case the OPU is out of radial tracking, (the radial distance different from zero), one of the photo detectors will see larger pit/land modulation than the other.

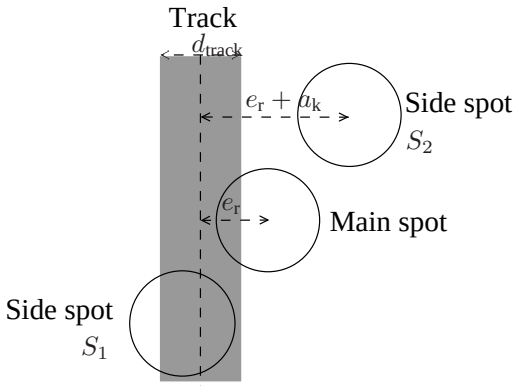


Figure 2.9: Illustration on how the three beams are positioned relative to each other on the disc surface. S_1 and S_2 are the two side detectors, d_{track} is the track width, e_{r} is the radial distance, and a_{k} is the offset of the side detectors.

Differential time detection principle The three beams principle is often used in CD-players, where the two other principles mentioned above are also used. However, they all have some severe disadvantages which disqualify them for use in DVD and HD drives. Instead another method is preferred called Differential Time Detection (DTD), or a variant called Differential Phase Detection (DPD). The idea is to detect the time or phase change of the high-frequency signals as a function of the radial distance of the scanning spot. These time or phase changes arise in a four quadrant photo detector. By processing the four signals detected in the four photo detectors the radial distance can be obtained, for more details on DPD/DTD see [Braat et al., 2002]. In practise the DTD is used, since the measuring of the differential phase change is more complicated.

2.7 Optical disc servos

Servo loops are in the optical disc players used to control three states: focus and radial distances and either the linear or angular speed of the OPU relative to the track. There is a number of different actuator configurations for the adjustment of focus and radial distances, see [Stan, 1998] and [Bouwhuis et al., 1985]. In the following only the configuration from the used CD-player is described. One servo is used to adjust focus distance. This servo is called the focus servo. Two servos are used to handle the radial corrections. The radial servo is used to do the fine adjustments of the radial distance and the larger adjustments are handled by the sledge servo. Finally one servo controls the disc speed.

Focus servo A linear electro-magnetic actuator performs the corrections in the focus distances, by positioning the moving lens in the focus direction. The control signal to the focus actuator is computed based on the focus distance measurement. Due to the non-linearity of the focus measurement, it is of high importance to keep the focus distance close to zero and at least in the linear area of the focus measurement mapping; for more details on the performance requirements see Section 2.8.1. The disturbances which move OPU out of focus have many different natures, see Section 2.8.2. It is very important to strongly suppress the disturbances, due to the small dimensions of the track, lands and pits. If the disturbance is not suppressed, the OPU will get out of focus and it is not possible to retrieve any information from the disc before the OPU has been focused again.

Radial servo The fine adjustments of the radial distances are performed in the same way as the focus adjustments, except that the radial linear electro-magnetic actuator moves the moving lens in the radial direction instead. The working range of the radial actuator is only a few hundred tracks, meaning that another servo is needed for the larger corrections in the radial directions, this servo is called the sledge servo, and will be described in the next paragraph. The focus and radial servos are designed to be decoupled, since the linear magnetic actuators are placed orthogonal to each other, however in practise cross-couplings can occur, see [Dettori, 2001] and [Yeh and Pan, 2000]. For single speed optical disc drives such as the CD-Audio, these cross-couplings can be neglected.

Sledge servo The sledge servo performs the larger radial corrections by moving the OPU in the radial direction, as illustrated in Fig. 2.4. The OPU is mounted on this sledge which is driven by motorised gears enabling the movements of the OPU in the entire program area of the optical disc. Ideally the sledge servo performed slow continuous adjustments while the radial servo handles the high frequent adjustments, meaning that the sledge servo do not introduce perturbations on the radial position, or in other words do disturb the radial servo. Unfortunately, this is not the case in practise, since where is a clear self-pollution in the radial servo originating from the sledge servo. Variations and imperfections in the sledge actuators can cause severe disturbances in the radial servo. In some cases of severe disturbances these can result in a loss of track situation. The hierarchy of the radial and sledge servos is often such that the sledge servo react on the radial control signal if it is too large. [Aangenent, 2002] and [Dzanovic and Lauritsen, 2003] did some dedicated work on the cross-coupling between the radial and sledge servos.

Disc servo The data in optical discs are recorded using a constant channel clock. As a consequence the disc servo should control the readout speed of the disc to be a Constant Linear Velocity (CLV) of 1.3 [m/s]). This results in a feeding of data to the data separator with a constant throughput at 150[Kb/s], also called 1xspeed. With CLV rotation frequency decreases from the lead in area to the lead out area (9[Hz]-3[Hz]). The rotation speed is in practise controlled by the use of fullness of a data buffer, which is

Conditions	Parameter	Requirement
Below 500[Hz]	Max. deviation	$\pm 500[\mu\text{m}]$
	Max. vertical acceleration	$10[\text{m/s}^2]$
Above 500 [Hz]	Max. deviation	$\pm 1.0[\mu\text{m}]$

Table 2.3: Standardised focus deviations from nominal position of the information layer, given the required scanning velocity $v_{\text{a}} = 1.2 \cdots 1.4$[m/s].

controlled not to under or overflow. If it is close to underflow the rotation frequency is increased and if it is close to overflow the rotation frequency is decreased. This controller is designed such that the CLV is in the interval between 1.2[m/s] to 1.4[m/s], see the Red Book [Philips and Sony Corporation, 1991]. In ROM drives where timing is not an issue, the disc is controlled to have a Constant Angular Velocity (CAV), see [Stan, 1998].

2.8 Performance, disturbances and defects

The design of the servos in the CD-player is obviously depending on: the performance requirements to the servos, disturbances which have an influence on the system, and finally defects and faults which cause changes in the CD-player and the CDs. The scope of this project has been limited to the focus and radial servos, meaning that only these servos are described in the following.

2.8.1 Performance requirements

It is clear that the OPU shall be focused and radially tracked at the information track on the disc, and if the OPU gets too much out of focus and/or tracking the OPU cannot retrieve the information. The Red Book [Philips and Sony Corporation, 1991] gives some requirements to the focus and radial servos in CD-players. The requirements guarantee that data readout is possible if the servos fulfill their performance requirements. Unfortunately, these requirements are based on CDs fulfilling the requirements to them. In practise many CDs do not fulfill these requirements. This means that the Red Book requirements can be viewed as absolute minimum requirements. The CD-player manufacturer has to obtain better servo performance than given in the Red Book, if their CD-players shall be able to play “all” CDs. The Red Book worst case focus and radial deviations, are shown respectively in Table 2.3 and Table 2.4. By assuming the deviation in focus and radial direction can be modelled as a harmonic function, and when differentiating it twice, an expression for the acceleration can be obtained. This is in [Vidal Sánchez, 2003] used to compute the maximum frequency at which the max. deviations are respectively $\pm 500[\mu\text{m}]$ and $\pm 70[\mu\text{m}]$. These frequencies are computed to be 22[Hz] and 12[Hz] respectively. The 500[Hz] have for the radial adjustment been

Conditions	Parameter	Requirement
Below 500[Hz]	Max. eccentricity of the tracking radius	$\pm 70[\mu\text{m}]$
	Max. radial acceleration	$0.4[\text{m/s}^2]$
Above 500 [Hz]	Max. track error	$\pm 0.03[\mu\text{m}]$

Table 2.4: Standardised radial deviations from nominal position of the information layer, given the required scanning velocity $v_{\text{a}} = 1.2 \cdots 1.4$[m/s]. The above 500[Hz] requirement is the measured radial error between 500[Hz] and 10[kHz] in closed loop with a controller. The open loop cross over frequency is 200[Hz] having an integration time of 0.02[s].

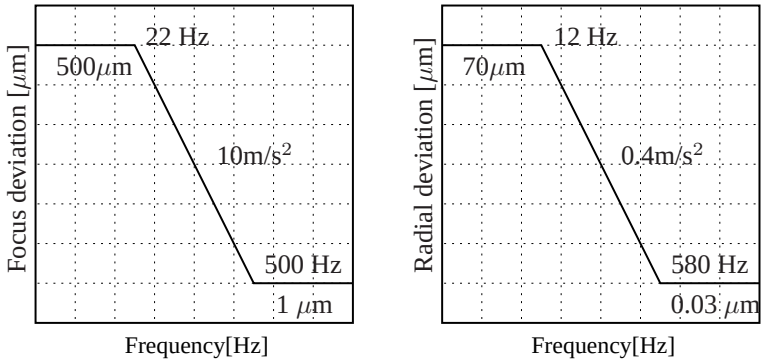


Figure 2.10: Graphic representation of the maximum focus and radial deviations on CDs for the required scanning velocity ($1.2 \ldots 1.4$ [m/s]).

moved to 580[Hz] due to the max acceleration requirement, see [Vidal Sánchez, 2003] and [Stan, 1998]. To clarify the requirements, the maximum focus and radial deviations are graphically represented in Fig. 2.10. In [Bouwhuis et al., 1985] approximations of the allowed maximal focus and radial deviations are given to be respectively: $\pm 2[\mu\text{m}]$ and $\pm 0.2[\mu\text{m}]$, due to non-linearity of the optical readout.

2.8.2 Disturbances and defects

Fig. 2.11 illustrates the main disturbances and defect influencing the focus and radial servos, both servos are in the figure merged into one multi-variable servo loop. The variables and names in Fig. 2.11 are defined as follow: Controller is the control, EMS is the electro-magnetic system in the OPU (both focus and radial), OS is the optical system which has the output of the photo detectors. $\mathbf{u}(t)$ is a vector of the control signals, $\mathbf{d}_{\text{m}}(t)$ is a vector of the mechanical disturbances, $\mathbf{x}(t)$ is a vector of the position of the OPU, $\mathbf{d}_{\text{d}}(t)$ is a vector of the disc deviations, $\mathbf{d}_{\text{s}}(t)$ is a vector of the self pollution, $\mathbf{e}(t)$ is a vector of the OPU position relative to the information track, $\mathbf{s}(t)$ is a vector of the defect

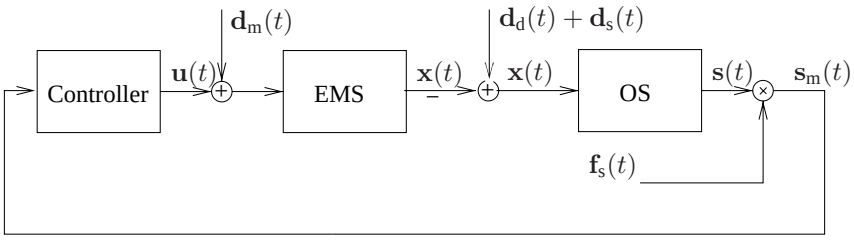


Figure 2.11: Illustration of the main disturbances and defects affecting the focus and radial loops. Controller is the controller, EMS is the electro-magnetic system in the OPU (both focus and radial), OS is the optical system which has the output of the photo detectors. $\mathbf{u}(t)$ is a vector of the control signals, $\mathbf{d}_\mathrm{m}(t)$ is a vector of the mechanical disturbances, $\mathbf{x}(t)$ is a vector of the position of the OPU, $\mathbf{d}_\mathrm{d}(t)$ is a vector of the disc deviations, $\mathbf{d}_\mathrm{s}(t)$ is a vector of the self pollution, $\mathbf{e}(t)$ is a vector of the OPU position relative to the information track, $\mathbf{s}(t)$ is a vector of the defect free photo detector outputs, $\mathbf{s}_\mathrm{m}(t)$ is a vector of the measured photo detector signals, and $\mathbf{f}_\mathrm{s}(t)$ is a vector of the surface faults.

free photo detector outputs, $\mathbf{s}_\mathrm{m}(t)$ is a vector of the measured photo detector signals, and $\mathbf{f}_\mathrm{s}(t)$ is a vector of the surface faults.
In order to handle all these disturbances and defects it is important to distinguish between disturbances which the controller shall suppress and defects/faults which the controller shall not react on. Disc deviations, mechanical disturbances and self pollution are disturbances. In the following the different groups will be described in more details.

Disturbances

- **Mechanical disturbances**, $d_\mathrm{m}(t)$: Originate from external sources, these disturbances propagate through the chassis of the CD-player, affecting the position of the OPU. The amplitudes of these disturbances depend on the CD-player and the environment in which it is used. Car-mounted CD-players and portable disc-mans are more exposed to mechanical disturbances than a stationary CD-player. The controller bandwidth in the two first applications are often higher than the controller bandwidth of the stationary CD-player. The higher bandwidth makes it possible for the controller to suppress these larger disturbances. Unfortunately the higher bandwidth has the drawback of being more sensible to surface defects. In addition to the use of controllers for suppressing the external disturbances, the OPU is mechanically decoupled from the chassis, this dampers works a low-pass on the disturbances applied to the chassis. The buffers in disc-mans and car mounted CD-players is often much larger than in stationary CD-players. These buffers are used to store the music and thereby ensure the constant stream of music, in spite of the mechanical disturbances.

- **Self pollution**, $d_{\mathrm{s}}(t)$: is caused by internal factors inside the CD-player, e.g. cross couplings between focus and radial actuators, sledge and radial, and disc servo to the other loops.

- **Disc deviations**, $d_{\mathrm{d}}(t)$: This group is due to deformations of the CD or CD-player in a way that cause position deviations of the information track. Some examples are: disc warp, skew discs and eccentric discs. These disturbances result in periodic disturbance with a natural frequency equal the rotation frequency of the disc. The maximal amplitudes, see [Philips and Sony Corporation, 1991], are given by the specifications in Tables 2.3 and 2.4. In addition inaccuracies in the mechanical clamping of the CD in the CD-player can also cause deviations in the position of the information track.

Defects

- **Surface defects**, $f_{\mathrm{s}}(t)$: The remaining challenges in control of CD-players are in this group. These surface defects change the read-out from the disc by imply changes on the disc surface and these surface defects are: scratches, fingerprints, dust and other surface debris, coating defects, birefringence due to molding problems, air bubbles, thickness irregularities, etc. The three defects: scratches, fingerprints and dust are of interest in the remainder of this thesis.

2.8.3 Surface faults

Scratches, fingerprints and dust are following denoted surface faults. The surface faults influence the photo detectors signal in such a way that the generated feedback signal also contains a faulty part due to the surface faults. During a surface fault the controllers will base their computation of the control signal $u(t)$ on the feedback signal containing a faulty component due to the surface fault. The controllers can force the OPU out of focus and / or radial tracking due to this fault feedback signal component. Almost all CD-players have implemented some kind of fault handling, it is often implemented by detecting the faults based on a fast drop in detected light energy. During a surface fault the feedback signals are not entirely reliable, this means that a simple way to handle them is to sample with a low enough sample frequency such that the surface fault can be placed between two samples, in practice this i done by decreasing the controller bandwidth during the occurrence of the surface fault. When the fault is detected the controller bandwidth is decreased, in practise the feedback signal is often fixed at zero, which introduces some unwanted controller reactions to this step in the feedback signal. At the detection of the end of the fault the bandwidth is put back to the normal, see [Philips, 1994], [Andersen et al., 2001] and [Vidal et al., 2001b]. However, this switch back to normal control does often cause more severe problems than the handling of the surface faults. The handling of both surface faults and mechanical disturbances is clearly conflicting, since the handling of surface fault requries a low bandwidth which

is in conflict with the requirement of a high bandwidth for handling disturbances. The fact that disturbances and surface faults are non separable in time domain and frequency domain also increase the complexity of solving this control problem. The error coding described in the Red Book allows faults of a size up to 2.5 [mm], and the used handling strategies do not handle anything near to that size. The handling of these surface faults is of the interest in the remainder of this thesis.

2.8.4 Playability

The Red Book defines the maximal physical tolerances of CDs, as well as gives specifications to the CD-players. These specifications ensure compatibility between the CDs and the CD-players. This means that it can be said that the playability is ensured as long as the Red Book requirements are fulfilled. In [Vidal Sánchez, 2003] a more broader definition is given as

Definition 2.1 (Playability) *the ability to reproduce the information from non-ideal discs in non-ideal circumstances.*

[Vidal Sánchez, 2003] also divides the playability constraints into three major groups: HF signal, data channel and servo. This Ph.D work focus on improvement of the playability by addressing the servo part. The servo part of the playability problem can again be divided into the disturbances and faults described in Section 2.8.2. The remainder of the thesis is focused on improving the playability regarding surface faults without lowering the playability towards the other servo issues. The work will especially be focused on the handling of surface faults originating from the use of the CDs, such as scratches and fingerprints, and not surface faults originating from the production of the CDs like black or white dots.

2.9 Proposed feature based control strategy

In order to improve the playability of discs with surface faults without lowering the playability of other servo issues, a fault tolerant control strategy seems to be a well suited choice. This is supported by the work in [Vidal Sánchez, 2003]. It might also be a good idea to use a joint time/frequency analysis to extract features used to detect and handle the surface faults. The general structure of the suggested strategy, called feature based control, is illustrated in Fig. 2.12.
The CD-player feeds the photo detectors signals to the residual generator, which computes fault residuals. These residuals and the distance measurements are following used to locate the faults and to extract features of the fault which can be used to accommodate the controller to handle the fault. The fault is first located in time, (also called fault detection), followed by the extraction of the useful features. Due to the small distance

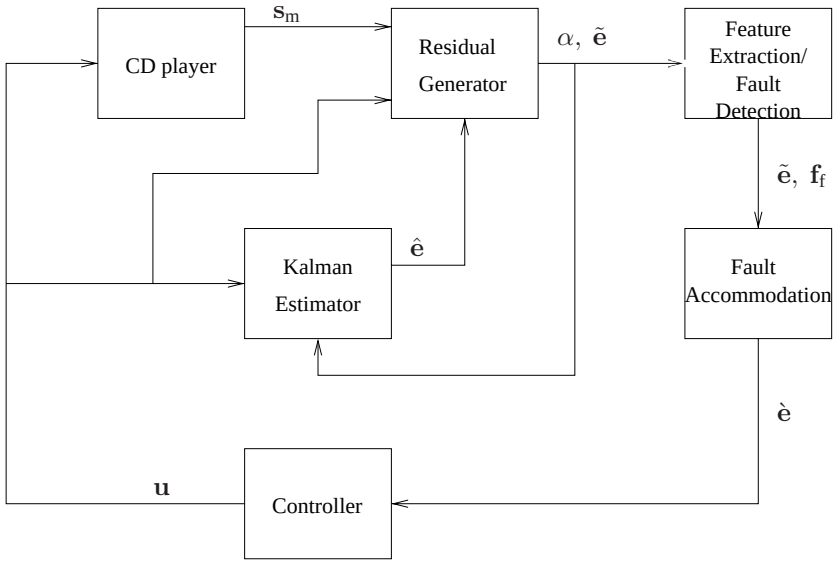


Figure 2.12: Illustration of the structure of the feature based control scheme, to handle surface faults as faults. $\mathbf{s}_m$ is vector of the measured detector signals, α is a vector of the residuals, $\tilde{\mathbf{e}}$ is a vector of the static estimates of the distances, $\mathbf{f}_f$ is a vector the fault features, $\hat{\mathbf{e}}$ is a vector of the dynamical estimates of the distances, $\grave{\mathbf{e}}$ is a vector of the corrected distances, $\mathbf{u}$ is a vector of the control signals.

between the tracks compared to normal scratches, these surface faults have small variations from one encounter to the next encounter, meaning that a repetitive principle can be used, see [Odgaard et al., 2004d]. This means that the feature at encounter M can be used to handle the fault at encounter $M + 1$, and when the fault is passed the feature extraction can be based on the entire fault, which clearly increases the potentials in this feature extraction. During the fault the extracted features are used to remove the fault component from the feedback signal, meaning that the controller reacts on a ideally fault free feedback signal.

2.10 Summary

In this chapter the general principles behind the optical disc players are described, these principles are used to address focus on the challenging control problems in optical disc players, with a special focus on the CD-player. Section 2.1 dealt with the different kinds and generations of optical disc players. The trend in the development of optical discs is to increase the information density, which leads to more challenging problems. The basic structure and dimensions of the optical disc were described in Section 2.2. In Section

2.3 the general structure of the optical disc players was revealed. Some required basic optical principles are described in Section 2.4, where these are used to reveal some of optical differences in the different generations of optical discs. This lead to a description of the optical pick-up, see Section 2.5. Due to the lack of physical contact between the optical pick-up and the information track on the optical disc, control is needed to focus and radially track the optical pick-up at the information track on the optical disc. Section 2.6 described the indirect measurement of the feedback signals used to position the optical pick-up - These feedback signals are used in servo loops to positioning the optical pick-up. The servo loops were the topic of Section 2.7. These servos are required to fulfill some performance requirements. The servo shall also suppress disturbances such as mechanical disturbances and disc deviations, and handle surface faults such as scratches by not reacting to these, see Section 2.8. These requirements are highly conflicting since disturbances and faults partly coexist in the same frequency and time region. A way to handle this control problem is to address it by a feature based fault tolerant control scheme. The control scheme is introduced in Section 2.9.

Chapter 3

Model of a Compact Disc Player

This chapter deals with the modelling of focus and radial actuators and sensors in the CD-player, these models are derived with the purpose of designing focus and radial controllers in a given CD-player. Normally these controllers in the servos are designed based on dynamical models of the mechanical and electro-magnetic parts of the CD-players. The optics are normally modelled by a simple linear static model, which is sufficient for the design of nominal controllers for the focus and radial servos. However, in the case of fault detection, there are some optical cross-couplings which makes early detection of the surface faults impossible if an optical model is not used.

A large effort is put in modelling and identification of the mechanical and electro-magnetic parts of CD-players. [Bouwhuis et al., 1985] focuses on the modelling of these parts of the system and [Vidal et al., 2001c] describes a simple method to perform open loop system identification. Both [Yeh and Pan, 2000] and [Dettori, 2001] perform some work on modelling the cross-couplings in the mechanical and electro-magnetic parts between focus and radial loop. Regarding the optical part of the system the present control strategies are based on simple linear models which do not concern the optical cross coupling [Bouwhuis et al., 1985] and [Stan, 1998], although some work has been done regarding the optical model. [Yeh and Pan, 2000] has some considerations about the optical cross-couplings. [Vidal et al., 2001b] and [Vidal et al., 2001a] deal with models of the optical signal of focus distance without consideration about the cross coupling with the radial loop. This model is used for detection of surface faults on the disc. The principle behind the optical pick-up applied in this work is the single Foucault three beam principle, which is described in [Stan, 1998], [Bouwhuis et al., 1985]. Based on these principles and measurements on a CD-player test setup, a model of a three beam single Foucault optical detector system is made. This model maps from focus and radial distances to focus sensors and radial sensor signals. The model includes the cross-couplings between focus and radial loops. This optical model is followed by a model of the surface faults.

The used experimental test setup is described in Section 3.1. In Section 3.2 a model

structure of the focus and radial servos is given, in Section 3.3 the mechanical and electro-magnetic model is derived. This leads to the main novel contributions in this chapter: the optical model in Section 3.4 and the surface fault model in Section 3.5.

3.1 Experimental test setup

Some of the results of this thesis were tested on an experimental test setup which will be described in this section, with a focus on the possibilities and limitations this experimental test system gives. A stereo music system with a CD-drive is used as main part of the experimental test system. The stereo system has implemented internal focus and radial PID controllers. These controllers are digitally implemented with a fixed size and structure. As a consequence it is only possible to change the parameters in these controllers. These limitations can be bypassed by using a PC-based experimental setup, as illustrated in Fig. 3.1. A photo of the experimental test setup is shown in Fig. 3.2. This experimental test setup is the same as used in [Vidal Sánchez, 2003], which again is based on a previous design documented in [Vidal et al., 1998] and [Andersen and Karlsson, 2000]. The original CD-drive is separated from the stereo music system, but still connected through the communication link such that the music stereo system can be operated as normal. The built-in PID controllers receive as inputs the four detector signals: $D_1,\ D_2,\ S_1,\ S_2$. The control signals computed by these PID controllers, u_{f} and u_{r} conducted through the PC-controlled switches to the CD-drive, which closes the control loops. An alternative to this setup is that the focus and radial positions are controlled by the controllers in the PC. The photo diode signals are fed through the PCI 9118-DG I/O-card to the PC. The photo diode signals are in this process amplified, low-pass filtered and discretised by the I/O-card. In the PC two control signals are computed and passed following through an attenuator stage. The PC also changes the PC-controlled switches so the CD-drive can be controlled by the PC instead of the built-in controllers. In addition the control signals computed by the internal PID controllers are also sampled by the PC.
In the following the relevant specifications of the experimental test setup are listed. A consequence of the relative old PC and I/O-card is the trade-off regarding the sample frequency and bandwidth of the anti-aliasing filter. Even though these are higher in industrial players, these are chosen due to computational limitations of the PC and I/O-card. However, the listed data are acceptable and usable trade-offs.

- Amplifier gain: Photo diode gain: 60 [dB], Control signal measurement gain: 40 [dB].
- Anti-aliasing filter: First order low pass filter with a 8 [kHz] bandwidth.
- Attenuation factor for control output: -54 [dB].
- PCI 9118-DG I/O-card: 8 12 bit $\pm$5 [V] analog inputs with on-chip sample-and-hold. 2 12 bit $\pm$10 [V] analog outputs. 35 [kHz] sample frequency for each

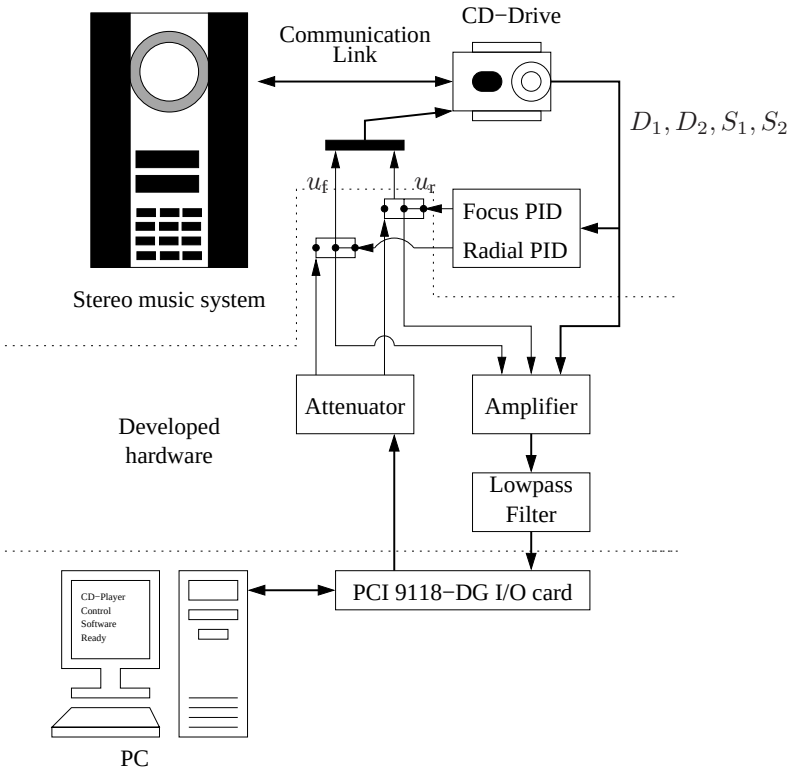


Figure 3.1: Overview of the experimental setup. The CD-drive of the stereo system is connected through the developed hardware to a PC. The focus and radial position can be controlled by either the internal PID controllers or by controllers implemented in the PC. One should notice that it is possible to control one position with the internal controller and the other position with the external controller.

channel. Direct Memory Access (DMA) data transfer.

- Personal Computer: AMD Athlon PC 1100 MHz with Dos 6.1 operating system installed and C-compiler.

The amplifier is the most critical part of the experimental setup. The OPU generates current signals below 12 [μA] which can not be sampled by the I/O-card. In addition signals with such a low signal power are very sensitive to measurement noises. Measurement noises is a problem due to a long wire between the OPU and hardware. Amplifiers are used to convert these signals to voltages in the range of the I/O-card.

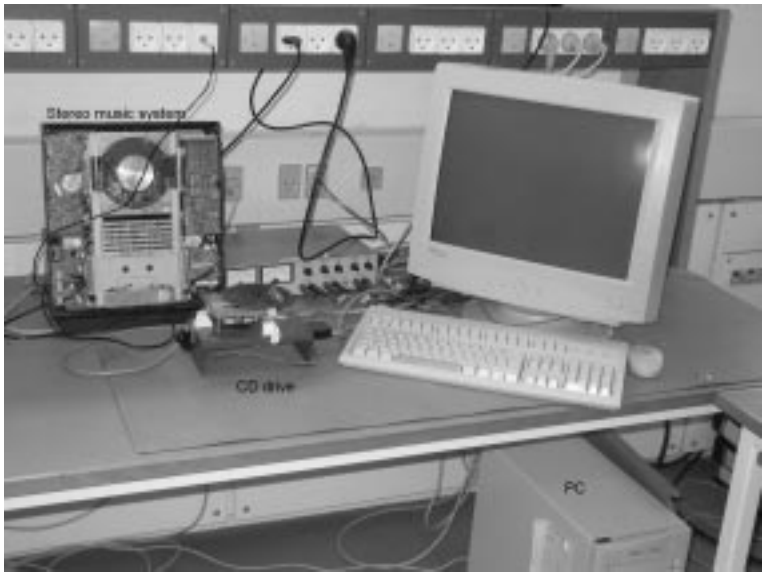


Figure 3.2: Photo of the experimental test setup.

CD-drive The CD-drive used in the experimental test setup is based on the following principles: A holographic optics arrangement, which implements the focus and radial optical detectors. The used focus detector principle is the single Foucault principle, and the used radial detector principle is the Three-beam method. The optics are positioned using a 2-axis device mounted on a sledge.
These detector and actuator principles are described in this chapter 2, and modelled later in this chapter.

Limitations An on-the-fly switch from the internal to the external controller is only possible in the focus loop. At the time of switching, any difference between the internal and external controller's output introduces a step on the control signals applied to the actuators. The focus loop accepts large errors ($\pm 6[\mu\text{m}]$) before loosing focus and has shown in practice to be robust towards these steps. The radial loop on the other hand has shown to be much more sensitive towards these steps. This is a consequence of the fact that a change larger than 0.8 μm results in a jump to the next track. The built-in logic in the stereo music system tries to correct that by jumping back to the previous revolution. However, this is not possible, since the radial position is controlled by the PC. The built-in logic will as a result generate a failure which stops the CD-player immediately. A way to work-around this problem is to implement the radial controller when the stereo music system is in test mode, where no actions are taken in case the presence of jump in tracks. It has the drawback that the sledge controller is not applied, meaning that the

radial controller can work only in a few seconds before it reaches the saturation limits. The low-budget PCI9118-DG I/O-card does also introduce some limitations in the test system. It is designed to be used in applications where data is sampled in a row and turn-around delay is not critical. In closed-loop systems these delays lead to instability and severe limitations in the performance. These delays should as a consequence be avoided. The used software on the PC was created in a way such that the delay is limited to two sample periods. A better alternative is used in [Dettori, 2001] where the experimental setup is based on dSpace, which offers a powerful design tool for control, but at a high price. In [Dettori, 2001] a hardware-software based solution where the radial controller can be changed on-the-fly with out the limitations of the experimental setup of the present Ph.D work.

3.2 Model structure

The lack of physical contact between the OPU and the information track on the CD, results in the need of focusing and radial tracking of the OPU. These are performed by the use of two servo loops, where the physical adjustments are performed by the 2-axis device, which enables positions adjustments in both the focus and radial direction. A number of alternative implementations of this devices are also used. Only the 2-axis device used in the test setup, in this work, will be described. This 2-axis device uses two linear electro-magnetic actuators to adjust the OPU in the focus and radial direction. These actuators are by design placed orthogonal to each other, resulting in theory in non-electro-magnetic cross-couplings between the two actuators. The OPU is suspended on four parallel arms, as illustrated in Fig. 3.4. The actuators enable the adjustments of the OPU, but are also necessary to measure the focus and radial distances in order to adjust them to be zero. By the use of smart optics, the OPU generates some indirect measurements of these distances. The indirect measurements can be performed in a number of ways. In this work the single Foucault method is used to measure the focus distance and the three beam principle is used to measure the radial distances. The main problem to solve in this Ph.D project is to handle surface faults that are impossible to handle by normally used methods. It is anticipated that the use of a model of surface faults would make this possible. This gives the third model block, enabling an illustration of the overall model structure. This model structure is illustrated in Fig. 3.3. The three sub-models are the electro-magnetic-mechanical model, the optical model and the surface fault model.

3.3 Electro-magnetic-mechanical model

The focus and radial actuators are in principle identical, meaning they can be described with the same model. As a consequence only the focus actuator will be described in the following. The electro-magnetic part of the OPU is again composed by two parts, a

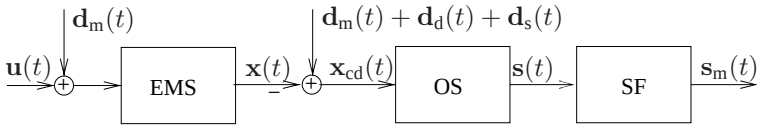


Figure 3.3: Illustration of general model structure of the combined focus and radial model. EMS is a model of the electro-magnetic-mechanical system in the OPU, OS is the optical model which has the output of the photo detectors, and SF is a model of the surface faults. $\mathbf{u}(t)$ is a vector of the control signals, $\mathbf{d}_\mathrm{m}(t)$ is a vector of mechanical disturbances, $\mathbf{x}(t)$ is a vector of the positions of the OPU, $\mathbf{d}_\mathrm{d}(t)$ is a vector of the disc deviations, $\mathbf{d}_\mathrm{s}(t)$ is a vector of the self pollutions, $\mathbf{x}_\mathrm{cd}(t)$ is a vector of the OPU positions relative to the information track, $\mathbf{s}(t)$ is a vector of the fault free photo detector outputs, $\mathbf{s}_\mathrm{m}(t)$ is a vector of the measured photo detector signals, and $\mathbf{f}_\mathrm{s}(t)$ is a vector of the surface faults.

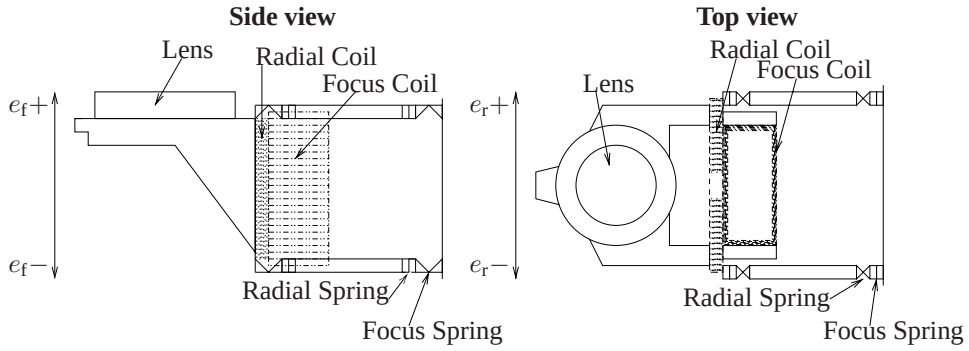


Figure 3.4: Illustration of 2-axis device suspended on four parallel arms, where two linear electro-magnetic actuators are used for the focus and radial adjustments.

driver part and a mechanical part. These two parts are described in more details in the following two subsections, ending up with a transfer function for the electro-magnetic model.

3.3.1 Electro-magnetic part

The driver performs the conversion from the applied control voltage $u(t)$ to the force acting on the mechanical part of the actuator $F(t)$. The active part in this actuator is the coil which is driven by a current. The power driver converts $u(t)$ to the current $i(t)$. The power driver is modelled by a constant gain K_pd and results in a different voltage across the linear electro-magnetic actuator, $u_\mathrm{coil}(t)$. The power driver is not physically placed in the OPU, but is considered to be a part of it, and is as a consequence a part of the driver part. A sketch of the driver part is drawn in Fig. 3.5.

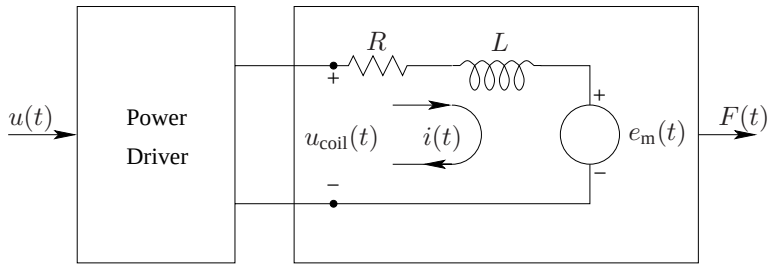


Figure 3.5: Sketch of the driver part of the model. The applied control voltage $u(t)$ is converted into a current $i(t)$. The current runs through the coil and in the magnetic field this results in a force $F(t)$. $u_{\text{coil}}(t)$ is the voltage over the coil, $e_{\text{m}}(t)$ is the induced voltage, R is the resistance in the coils, and L is the inductance in the coils.

This linear electro-magnetic actuator behaves like a direct-current (dc) motor for restricted displacements, which is the case of the normal use of the actuator. The dc motor equivalent electrical diagram is shown in Fig. 3.5. The equivalent is composed by a resistance R, an inductance L and the induced voltage source $e_{\text{m}}(t)$. By using Kirchhoff's voltage law $u_{\text{coil}}(t)$ can be computed

$$u(t) \cdot K_{\text{pd}} = u_{\text{coil}}(t) = R \cdot i(t) + L \cdot \dot{i}(t) + e_{\text{m}}(t). \tag{3.1}$$

The induced voltage $e_{\text{m}}(t)$ is a result of the coil movements relative to the magnetic field. $e_{\text{m}}(t)$ is described by the following equation

$$e_{\text{m}}(t) = (\dot{\mathbf{x}}(t) \times \mathbf{B}) \cdot \mathbf{l}, \tag{3.2}$$

where $\dot{\mathbf{x}}(t)$ is the velocity of the coil in the magnetic field, $\mathbf{B}$ is the flux density of the magnetic field and l is the length of the moving coil. In practice this cross product can be replaced by a scalar product since the three vectors are mutually orthogonal. This results in a simplification of (3.2)

$$e_{\text{m}}(t) = \mathbf{B} \cdot \mathbf{l} \cdot \dot{\mathbf{x}}(t). \tag{3.3}$$

$i(t)$ can be calculated by substituting e_{m} from (3.3) into (3.1), followed by an isolation of $i(t)$

$$i(t) = \frac{u(t) \cdot K_{\text{pd}} - B \cdot l \cdot \dot{x}(t)}{R + L \cdot s}. \tag{3.4}$$

A different physical law governs the electrical induced force, $F(t)$. The conductor is assumed to be fixed for simplicity. The current through the conductor in interactions with the magnetic field generates a force on the conductor. It can be described by the following equation

$$\mathbf{F}(t) = i(t) \cdot (\mathbf{B} \times \mathbf{l}). \tag{3.5}$$

Again the vectors are mutually orthogonal, meaning that (3.5) can be simplified to

$$F(t) = B \cdot i(t) \cdot l. \tag{3.6}$$

By assuming the system to be linear a transfer function in the laplace domain can be stated, by laplace transform (3.4) into (3.6). $F(s)$ can now be calculated by substituting $i(s)$ in the two laplace transformed equations.

$$F(s) = B \cdot l \cdot \frac{u(s) \cdot K_{\text{pd}} - B \cdot l \cdot s \cdot x(s)}{R + l \cdot s}. \tag{3.7}$$

3.3.2 Mechanical part

The force generated by the electro-magnetic part, $F(t)$ is applied to the OPU of the mass m. A simple model of the mechanical part is suggested in [Bouwhuis et al., 1985]. Assume the effects from friction, elasticity, etc. to be neglectable, then this system will act like a double integrator, described by Newton's second law, $F = m\ddot{x}$, where $\ddot{x}$ is the acceleration of the OPU. The position follows by integrating the acceleration twice. The mechanical parts are assumed to be linear, and it is whereby possible to state the force to position transfer function in the Laplace domain given as

$$\frac{x(s)}{F(s)} = \frac{1}{m \cdot s^2}. \tag{3.8}$$

In principles all optical disc players have mechanical parts which can be modelled as being double integrators, see [Bouwhuis et al., 1985]. However, the OPU is not floating as assumed in (3.8), it is suspended on leaf-springs. Taking the these leaf-springs into account the mechanical part can be modelled by a mass-spring-damper model, as illustrated in Fig. 3.6. m is the mass of the moving OPU, $x(s)$ is the displacement, b and k is respectively the damping and spring constant, and $F(s)$ is the applied force. This gives a new transfer function

$$\frac{x(s)}{F(s)} = \frac{1}{m \cdot s^2 + b \cdot s + k}. \tag{3.9}$$

This model structure is used in this work, even though a more precise model would be to introduce an parasite mass-spring-damper system. The pole which the additional mass-spring-damper introduces is placed at a high frequency where this work does not need a precise model. This means that the transfer function in (3.9) is used as a model of the mechanical part of the system.

3.3.3 General electro-magnetic model

This model can be achieved if the models of the mechanical part and the electro-magnetic part are merged. In order to achieve a transfer function in the Laplace domain, the electro-magnetic part of this model is assumed to be linear. The transfer function

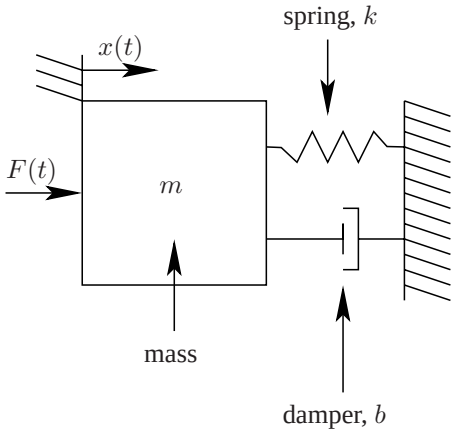


Figure 3.6: Free body diagram of the mass-spring-damper system.

from the applied focus control voltage $u_{\mathrm{f}}(s)$ to the focus distance $e_{\mathrm{f}}(s)$ can be found by substituting the force from (3.7) into (3.9), and followed by an isolation of the required variables. The transfer function is derived as

$$\frac{e_{\mathrm{f}}(s)}{u_{\mathrm{f}}(s)} = K_{\mathrm{pd}} \cdot \frac{\frac{B \cdot l}{m \cdot R}}{\frac{L}{R} \cdot s \cdot \left(s^2 + \frac{b}{m} \cdot s + \frac{k}{m}\right) + \left(s^2 + \left(\frac{(B \cdot l)^2}{m \cdot R} + \frac{b}{m}\right) \cdot s + \frac{k}{m}\right)}. \quad (3.10)$$

The radial transfer function from $u_{\mathrm{r}}(s)$ to the radial distance $e_{\mathrm{r}}(s)$, can be derived in the same way. The test-setup used is the same as used in [Vidal Sánchez, 2003]. The pole related to the inductance L is in [Vidal Sánchez, 2003] approximated to above 10[kHz] for both the focus and radial actuators, which are above the typical closed loop bandwidth of both the focus and radial loops in CD-players, since the interval of typical closed loop bandwidth is 0.5-2[kHz], (see for example [Steinbuch et al., 1994], [Yokoyama et al., 1994], [Chait et al., 1994], [Huang et al., 1999], [Moon et al., 1996] and [Dettori et al., 1999]). This means that the inductor can be neglected.

Model adapted to the test-setup drive By neglecting the pole due to the inductance, both the focus and radial actuators can be described with second order transfer functions. The focus transfer function is

$$\frac{e_{\mathrm{f}}(s)}{u_{\mathrm{f}}(s)} = K_{\mathrm{pd,f}} \cdot \frac{\frac{B_{\mathrm{f}} \cdot l_{\mathrm{f}}}{R_{\mathrm{f}} \cdot m_{\mathrm{f}}}}{s^2 + \left(\frac{b_{\mathrm{f}}}{m_{\mathrm{f}}} + \frac{(B_{\mathrm{f}} \cdot l_{\mathrm{f}})^2}{m_{\mathrm{f}} \cdot R_{\mathrm{f}}}\right) \cdot s + \frac{k_{\mathrm{f}}}{m_{\mathrm{f}}}}, \quad (3.11)$$

where: $K_{\mathrm{pd,f}}$ is the focus power driver gain, B_{f} is the magnetic flux density in the focus actuator, l_{f} is the length of the moving coil in the focus actuator, R_{f} is the resistance in

the focus actuator, m_{f} is the mass which is moved by the focus actuator, k_{f} is the spring constant of the focus actuator and b_{f} is the damper constant of the focus actuator. The radial transfer function is

$$\frac{e_{\mathrm{r}}(s)}{u_{\mathrm{r}}(s)} = K_{\mathrm{pd,r}} \cdot \frac{\frac{B_{\mathrm{r}} \cdot l_{\mathrm{r}}}{R_{\mathrm{r}} \cdot m_{\mathrm{r}}}}{s^2 + \left(\frac{b_{\mathrm{r}}}{m_{\mathrm{r}}} + \frac{(B_{\mathrm{r}} \cdot l_{\mathrm{r}})^2}{m_{\mathrm{r}} \cdot R_{\mathrm{r}}} \right) \cdot s + \frac{k_{\mathrm{r}}}{m_{\mathrm{r}}}}, \tag{3.12}$$

where: $K_{\mathrm{pd,r}}$ is the radial power driver gain, B_{r} is the magnetic flux density in the radial actuator, l_{r} is the length of the moving coil in the radial actuator, R_{r} is the resistance in the radial actuator, m_{r} is the mass which is moved by the radial actuator, k_{r} is the spring constant of the radial actuator and b_{r} is the damper constant of the radial actuator. In [Vidal Sánchez, 2003] these model parameters are estimated, and since Vidal performed this parameters estimation on the same test setup as used in this work, the parameters estimated in [Vidal Sánchez, 2003] are used. The models can be seen in Appendix A.

3.4 Optical model

The often used model of the optical systems of the CD-player is a simple model, assuming the focus and radial distances to be linearly dependent on the difference on their respectively photo detectors, see [Bouwhuis et al., 1985], [Stan, 1998] and [Pohlmann, 1992]. Doing so one neglects the optical cross-coupling and the non-linear behaviour of the optical system. This optical cross-coupling is not so important in the estimations of the focus and radial distances. It is much more important in the detection of surface faults. The normal strategy is to detect fast changes in the received energy, see [Andersen et al., 2001] and [Philips, 1994]. The amount of received energy can approximated by the sum of the two focus detectors or the two radial detectors. Unfortunately, the optical cross-coupling has a clear contribution to these sum signals, meaning that the focus sum depends on the radial distance, and the radial sum depends on the focus distance. This indicates that a model of the optical system mapping from focus and radial distances to the four detector signal would be of good use.
The modelling of CD-players have been focused on the electro-magnetic parts of the CD-player, although some work has been done regarding the optical model. [Yeh and Pan, 2000] have some considerations about the optical cross-couplings. [Vidal et al., 2001b] and [Vidal et al., 2001a] deal with models of the optical signal of the focus error without consideration about the cross-couplings with the radial loop. This model is used for detection of surface faults on the disc. In DVD-players [Filardi, 2003] and [Hniličkа et al., 2002] have modelled the focus and radial mappings without taking any cross-couplings into account. Focus has been put on the optical model of the read-out process, see [Honguh, 2002] and [Kojima and Wakabayashi, 1997], which is not the same as the modelling of the optical detectors, even though some similarities do occur. The optical model of the CD-player used in this Ph.D project can be found in [Odgaard et al., 2003c], which models a CD-player with three beam single Foucault

detectors. The optical model is a static model, and the variables are in the following mentioned without a time dependence.

3.4.1 The three beam single Foucault detector principle

In the CD-drive, used in this work, the main beam is used both to restore the information saved on the disc, and to focus the beam at the disc surface. Two additional beams are used to keep the main beam radially tracked. In the following the focus and radial tracking detector principles are shortly described.

The focus detector The focus detector consists of two detectors, D_1 and D_2. The idea is to introduce some asymmetry in the light path from the disc surface to these detectors in such a way, that if the light beam is focused $D_1 - D_2 = 0$, and if the pick up is too far away from the disc $D_1 - D_2 > 0$, and if it is too close $D_1 - D_2 < 0$. This asymmetry can be generated in a number of ways, e.g. by the single Foucault principle which is illustrated in Fig. 3.7. The idea behind this focusing principle is to place a knife into

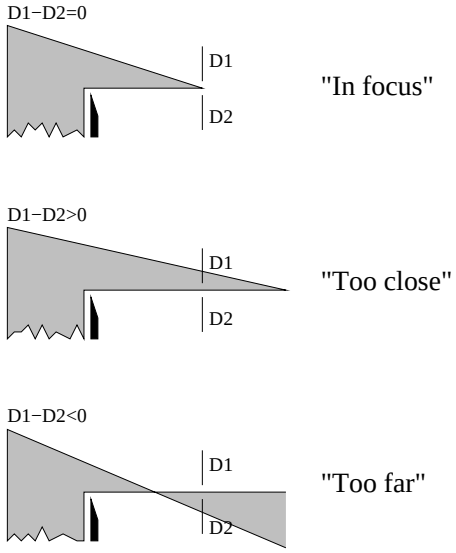


Figure 3.7: Illustration of the single Foucault focus detector principle. Three situations are illustrated: pick-up focused, pick-up too close to the disc, pick-up too far away from the disc.

the light path, such that only half of the light beam passes the knife, and the remainder is absorbed by the knife. The optical system is designed in such a way that if the light

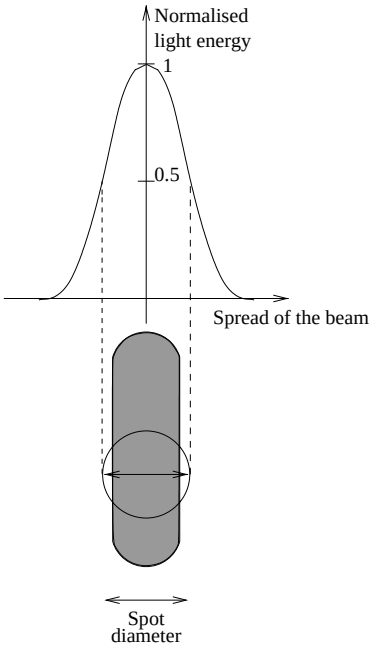


Figure 3.8: Illustration of the light intensity, and the approximation by assuming the light source being a disc with the radius equal to the radius of the half-intensity level.

beam is focused on the disc it would also be focused on the detectors. In the cases where the pick-up is either too close to or too far away from the disc, the beam focus point would be either behind or in front of the detectors. The principle illustrated in Fig. 3.7 is based on the assumption that the light source is point source. A better description of the source, however, is to use the energy distribution of the laser beam which is shaped like a sinc function, see [Stan, 1998] and [Bouwhuis et al., 1985], as illustrated in Fig. 3.8. This is simplified by assuming the light source being a disc with a radius equal to the radius of the half-intensity level, and with a uniform light intensity.

The disc shaped light source results in light being detected on both D_1 and D_2 in almost all situations, but the principle is the same as illustrated in Fig. 3.7. The sign of the difference between D_1 and D_2 is the same as in Fig. 3.7.

Radial detector Fig. 3.9 illustrates how the three beams, (main beam and the two side beams S_1 and S_2), are placed relative to each other on the disc surface. The main spot is placed in the middle of the track and the two others placed one to each side of the track, with a distance from their centre to the middle of the track called a_{k}. If the pick-up is located symmetrically over the track, the two side spots will cover equal areas of the

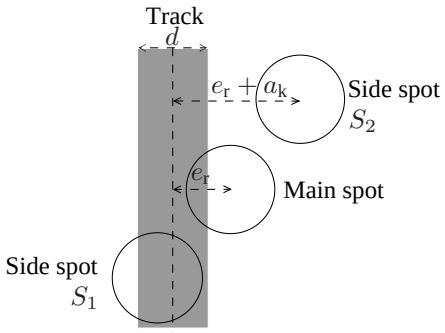


Figure 3.9: Illustration on how the three beams are placed relative to each other on the disc surface. In this situation the main spot is out of radial track with the distance e_r.

track, due to the fact that the disc spins around and that the detector signals are low-pass filtered. This means that the radial detector signals do not depend on the individual pits and lands. Since the low-pass filtering results in radial detector signals being the mean of pits and lands. Fig. 3.9 illustrates a situation where the pick-up is too much to right. In this case S_2 receives more light energy than S_1, due to the destructive interferences. This means that $S_1 - S_2 < 0$ when the pick-up is too much to the right. On the other hand if the pick up is too much to the left, the difference $S_1 - S_2 > 0$.

3.4.2 The focus and radial models

This optical model is expressed by the mappings, described by (3.13-3.16).

$$f_1 : (e_f, e_r) \rightarrow D_1, \tag{3.13}$$
$$f_2 : (e_f, e_r) \rightarrow D_2, \tag{3.14}$$
$$f_3 : (e_f, e_r) \rightarrow S_1, \tag{3.15}$$
$$f_4 : (e_f, e_r) \rightarrow S_2. \tag{3.16}$$

These mappings can be combined into one vector function, and if a vector containing the photo detector signals is named $\mathbf{s}$, the model can be written as

$$\mathbf{s} = \begin{bmatrix} D_1 \\ D_2 \\ S_1 \\ S_2 \end{bmatrix} = \mathbf{f}(e_f, e_r). \tag{3.17}$$

Due to limitation in the test setup it is only possible to verify and identify parameters in the mappings if the following assumption about these mapping is used.

$$f_i(e_\mathrm{f}, e_\mathrm{r}) \approx h_i(e_\mathrm{f}) \cdot g_i(e_\mathrm{r}), \tag{3.18}$$

where

$$i \in \{1, 2, 3, 4\}. \tag{3.19}$$

This assumption means that the radii of the spots at the disc is constant, even though the radii of the spots depend on e_f. However, an increase in the actual area of the spot is not important for this model. The important thing is the ratio between the area of the spot covering the track, A_{track}, and the area of the part of the spot covering some of the area between the tracks, A_{CD}. This ratio is almost constant. This means that influence of e_f on $g(\cdot)$ can be neglected as a consequence, see [Odgaard et al., 2003c].
(3.18) can be interpreted as follows. The $g_i(e_\mathrm{r})$ function computes the maximal detected energy for a given value of e_r. The $h_i(e_\mathrm{f})$ computes how large a ratio of the reflected energy in the spot which is detected at the photo detectors. In the following expressions for the $h_i(e_\mathrm{f})$ and $g_i(e_\mathrm{r})$ functions are derived.

The $h_i(e_\mathbf{f})$ mappings will be derived in the following, starting with $h_1(e_\mathrm{f})$ and $h_2(e_\mathrm{f})$. These two mappings are related in the following way:

$$h_1(e_\mathrm{f}) = h_2(-e_\mathrm{f}), \tag{3.20}$$

due to the linear movement of the focus point relative to the detectors. As a consequence of (3.20) it is only needed to model one of these function, $h_1(e_\mathrm{f})$ is modelled. This model of how e_f influences the diode signals, consists of two factors, $h_1(e_\mathrm{f}) = h_1^{'}(e_\mathrm{f}) \cdot \eta_1(e_\mathrm{f})$. The first factor, $h_1^{'}(e_\mathrm{f})$, is due to the single Foucault principle. The light beam at the detector would have some kind of half disc shape, due to the single Foucault principle. The radius of the half disc, r, increases linearly with e_f. Factor number two, $\eta_1(e_\mathrm{f})$, is the ratio of light energy detected at the detectors compared to the amount of light energy intended to be detected. The ratio decreases, since the detectors have to be relatively small due to the implementation and in order to minimise the noise received through the detectors. As a consequence the light beam covers more and more area outside the detector as e_f increases.
The modelling of the optical detectors can now be started with, $h_1^{'}(e_\mathrm{f})$, the first factor in the model. First all the lenses in the light path from light source to the disc surface are merged into one lens, F. This simplification is illustrated in Fig. 3.10
The distance from the lens to the disc surface and back to the lens again, is denoted l_x:

$$l_\mathrm{x} = l_{\mathrm{x},0} - 2 \cdot e_\mathrm{f}. \tag{3.21}$$

The distance from the source to the lens, l_u is:

$$l_\mathrm{u} = l_{\mathrm{u},0} + e_\mathrm{f}, \tag{3.22}$$

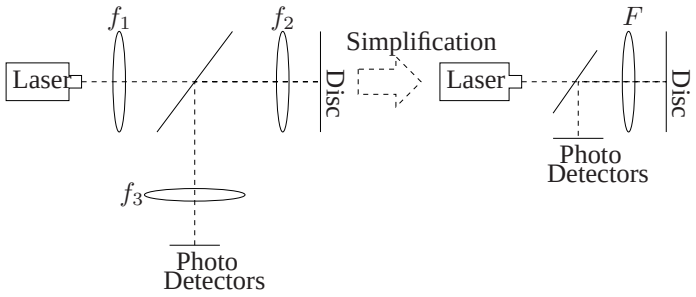


Figure 3.10: The optical path with three lenses ($f_1,\ f_2,\ f_3$), and the simplified version where the three lenses are approximated by one lens, F.

where: $l_{\mathrm{x},0}$ and $l_{\mathrm{x},0}$ are constants. The light source and detector are placed at almost the same place so the light travels through the same lenses F both ways. This means that by using the rules of thin lenses the focus point near the source and detector, F_1 can be found by:

$$F_1 = \frac{1}{\frac{1}{F} - \frac{2}{l_{\mathrm{x}}}} \tag{3.23}$$

The next step is to find how large a ratio of the energy sent from the source which is detected at the photo detectors. The disc shaped source can be divided into lines with a given vertical coordinate h_{op}. Fig. 3.12 illustrates how much of the energy emitted from a point in the line which is received at a point in the line at the detector with the vertical coordinate y_{op}. The figure illustrates that light travelling in the gray area passed through to the detector. The area of the beam at the lens what passes through is called A_{l}, the light spot at the lens has a radius R_{op}, and the ratio of light detected relative to the sent light is: $\frac{A_{\mathrm{l}}}{\pi \cdot R_{\mathrm{op}}^2}$, and A_{l} is

$$A_{\mathrm{l}} = \frac{1}{2} \cdot \pi \cdot R_{\mathrm{op}}^2 + R_{\mathrm{op}}^2 \cdot \arcsin(\frac{x}{R_{\mathrm{op}}}) + x \cdot \sqrt{R_{\mathrm{op}}^2 - x^2}, \tag{3.24}$$

where

$$x = \frac{k \cdot h}{l_{\mathrm{m}} - k} \cdot \frac{F_1}{l_{\mathrm{x}} - F_1}, \tag{3.25}$$

$$k = l_{\mathrm{u}} - k_0, \tag{3.26}$$

$$l_{\mathrm{m}} = \frac{l_{\mathrm{u}}}{l_{\mathrm{u}} - F_1} \cdot F_1. \tag{3.27}$$

By integrating the ratio over the half disc, ($[0; R_{\mathrm{source}}]$), where R_{source} is the radius of the source, $h_1^{'}(e_{\mathrm{f}})$ are found. This integration is done numerically by splitting the half disc

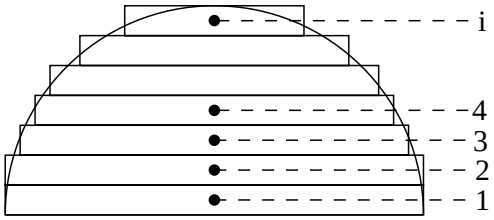


Figure 3.11: Illustration on how the half disc approximated with rectangles with centres in x_i and with the areas a_i.

up into approximation rectangles, and summing these. Fig. 3.11 is an illustration of the approximation of the half disc with rectangles. The height of the rectangles, h_{op} is given by:

$$h_{\text{op}} = \frac{R_{\text{source}}}{P_{\text{op}}}, \tag{3.28}$$

where P_{op} is the number of rectangles used for the approximation. These thin rectangles tend to a line as P_{op} increases and thereby approximates the lines with a constant vertical coordinate x. This method results in P_{op} x values, i.e. $x_i, i \in [1, 2, \cdots, P_{\text{op}}]$. A relative area function is defined as a function of i, $A_{\text{rel}}(i)$. This relative area tells how large a part of the energy in the laser which is in a given vertical line. It is now possible to approximate the integration over all the lines a given vertical constant, by summarising all the $A_{\text{l}}(x_i)$ multiplied with $A_{\text{rel}}(i)$:

$$h'_1(e_{\text{f}}) = \frac{1}{2} + \sum_{0}^{P_{\text{op}}} \frac{R_{\text{op}}^2 \arcsin\left(\frac{x_i}{R_{\text{op}}}\right) + x_i \cdot \sqrt{R_{\text{op}}^2 - x_i^2}}{\pi \cdot R_{\text{op}}^2} \cdot A_{\text{rel}}(i). \tag{3.29}$$

The modelling of $\eta_1(e_{\text{f}})$ is basically to find the ratio of the light energy detected at the detector relative to the energy intended to be detected. The shape of the detector is assumed to be a rectangle ($b_{\text{op}} \times 2b_{\text{op}}$). Fig. 3.13 illustrates the three possible situations on how the reflected spot covers the detector. By inspecting Fig. 3.13, the expression of $\eta_1(e_{\text{f}})$ can be found to be:

$$\eta_1(e_{\text{f}}) = \begin{cases} 1 & \text{if a)} \\ \frac{\frac{1}{2}\cdot\pi r(e_{\text{f}})^2 - r(e_{\text{f}})^2\cdot(2\theta_1(e_{\text{f}}) - \sin(2\cdot\theta_1(e_{\text{f}})))}{\frac{1}{2}\cdot\pi r(e_{\text{f}})^2} & \text{if b)} \\ \frac{2\cdot b^2}{\frac{1}{2}\cdot\pi\cdot r(e_{\text{f}})^2} & \text{if c)} \end{cases}, \tag{3.30}$$

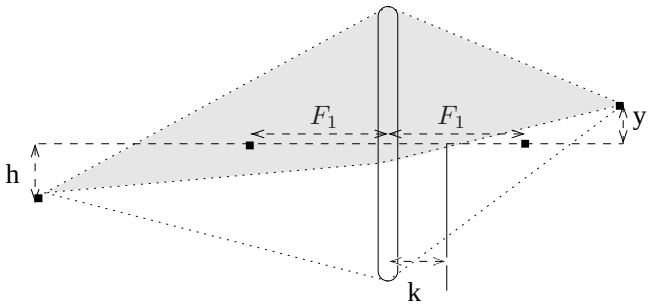


Figure 3.12: Illustration on how much light emitted from the detector at height h_{op} is absorbed by the knife, and the remainder received at height y_{op} at the detectors. The gray area of the beam is passed through to the detectors.

where

$$\theta_1(e_{\text{f}}) = \arccos(b_{\text{op}}/r(e_{\text{f}})), \tag{3.31}$$

$$r(e_{\text{f}}) = r_0 + \alpha \cdot e_{\text{f}}\,, \tag{3.32}$$

$$r_0, \text{ are two parameters,} \tag{3.33}$$

$$\text{a. } r < b_{\text{op}},$$

$$\text{b. } b_{\text{op}} \leq r \leq \sqrt{2 \cdot b_{\text{op}}^2},$$

$$\text{c. } r > \sqrt{2 \cdot b_{\text{op}}^2}.$$

$h_1(e_{\text{f}})$ can now be determined as the product of $h_1'(e_{\text{f}})$ and $\eta_1(e_{\text{f}})$:

$$h_1(e_{\text{f}}) = h_1'(e_{\text{f}}) \cdot \eta_1(e_{\text{f}}). \tag{3.34}$$

Due to the symmetry in (3.20), $h_2(e_{\text{f}})$ can be determined based on (3.34)

$$h_2(e_{\text{f}}) = h_1'(-e_{\text{f}}) \cdot \eta_1(-e_{\text{f}}), \tag{3.35}$$

$h_1(e_{\text{f}})$, (3.34), and $h_2(e_{\text{f}})$, (3.35), are now derived.
The next step is to find $h_3(e_{\text{f}})$ and $h_4(e_{\text{f}})$, and can be determined based on similar principles. The side beams follow the same light path through the lenses as the main beam. However, the single Foucault effect is not applied to the side detectors. Instead e_{f} influences the radial detector signals by the radius of the beam disc at the detector. As e_{f} increases more and more of the main spot is detected at the side detectors. This relationship is highly dependent on how the detectors are placed to each other, e.g. what the distance between the detectors is, etc. Due to this it is difficult to model the ratio

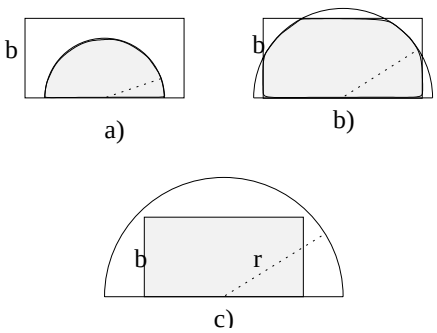


Figure 3.13: Illustration on how the reflected light covers the detector and area outside the detector as radius of the reflected half disc beam increases. a) $r < b_{\text{op}}$, b) $b_{\text{op}} \leq r \leq \sqrt{2 \cdot b_{\text{op}}^2}$ and c) $r > \sqrt{2 \cdot b_{\text{op}}^2}$

of the energy from the main spot is placed at the side detectors as e_{f} increases. Instead this part is approximated by a polynomial as a function of e_{f}, meaning $h'_{3,4}(e_{\text{f}}) = p(e_{\text{f}})$. Based on data this effect seems to dominate while the side spots are inside their detectors, they seems to grow out of there detectors at larger focus distance than for the two focus detectors. However, when the side spots start to grow out of their detectors the $\eta_i(e_{\text{f}})$-function, see (3.30), this effect is clearly the main contribution to mapping. These two model cases are merged into one model, by using the polynomial for small values of e_{f} and using the other part for larger values of e_{f}. In Section 3.4.3 the parameters in the models are identified based on measurements. The expression is as follows:

$$h_3(e_{\text{f}}) = \begin{cases} p(e_{\text{f}}) & \text{if a)} \\ \frac{\frac{1}{2}\cdot\pi r(e_{\text{f}})^2 - r(e_{\text{f}})^2\cdot(2\theta_1(e_{\text{f}}) - \sin(2\cdot\theta_1(e_{\text{f}})))}{\frac{1}{2}\cdot\pi r(e_{\text{f}})^2} \cdot k_{\text{s}} & \text{if b)} \\ \frac{2\cdot b^2}{\frac{1}{2}\cdot\pi\cdot r(e_{\text{f}})^2} \cdot k_{\text{s}} & \text{if c)} \end{cases}, \tag{3.36}$$

where

$$\theta_1(e_{\text{f}}) = \arccos(b_{\text{s}}/r(e_{\text{f}})), \tag{3.37}$$

$$r(e_{\text{f}}) = r_0 + \alpha \cdot e_{\text{f}}, \tag{3.38}$$

$$p(e_{\text{f}}) \text{ is a polynomial function of } e_{\text{f}}, \tag{3.39}$$

$$k_{\text{s}} = p(\frac{b_{\text{s}} - r_0}{\alpha}), \tag{3.40}$$

$$b_{\text{s}} \text{ is the height of the detector.} \tag{3.41}$$

$$\tag{3.42}$$

$$\text{a). } r < b,$$
$$\text{b). } b \leq r \leq \sqrt{2 \cdot b^2},$$
$$\text{c). } r > \sqrt{2 \cdot b^2}.$$

$g_i(e_\mathbf{r})$**:** These functions model the ratio of light energy in each of the light beams which are reflected at the disc. The functions $g_1(e_\mathrm{r})$ and $g_2(e_\mathrm{r})$ relates both to the main spot, i.e. $g_1(e_\mathrm{r}) = g_2(e_\mathrm{r})$. Two extra detectors are used to detect the side beams, i.e. $g_1(e_\mathrm{r}) \neq g_3(e_\mathrm{r}) \neq g_4(e_\mathrm{r})$.
The principle in this model is basically that, when the spot is moved over the track $A_{\mathrm{CD}}(e_\mathrm{r})$ and $A_{\mathrm{track}}(e_\mathrm{r})$ change depending on e_r. By changing e_r in Fig. 3.9 all three spots move, and they cover the track and area surrounding the track in different ratios depending on the radial distance. This means that the model of these three spots are the same except from an offset, a_k. So in the following only the main spot is modelled, $g_1(e_\mathrm{r})$, since $g_3(e_\mathrm{r}) = g_1(e_\mathrm{r} - a_\mathrm{k})$ and $g_4(e_\mathrm{r}) = g_1(e_\mathrm{r} + a_\mathrm{k})$. The track consists of both pits and lands. But no specific pits and lands are seen in the photo detector signal due to the low-pass filtering of these signals. Instead the detector signals can be modelled as the mean of the reflections from the pits and lands which the spots cover as the disc spins around. An approximation of this can be achieved by taking the mean of the two situations which occur when the spot covers some of the track. The spot either covers some of a pit or a land. $g_1(e_\mathrm{r})$ can be expressed as

$$g_1(e_\mathrm{r}) = \frac{1}{2} \cdot \left(\left(1 + \frac{A_{\mathrm{track}}(e_\mathrm{r})}{A}\right) \cdot \rho_{\mathrm{land}}\right) \cdot E_{\mathrm{beam}} - \frac{1}{2} \cdot \left(\frac{A_{\mathrm{track}}(e_\mathrm{r})}{A} \cdot \rho_{\mathrm{pit}}\right) \cdot E_{\mathrm{beam}} \quad (3.43)$$

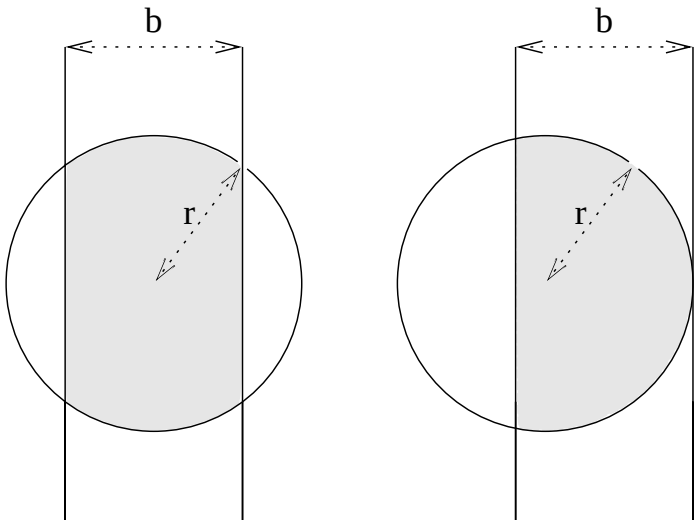
where: ρ_{land} and ρ_{pit} are the reflection ratios of a land and a pit. E_{beam} is the energy in the beam before it is reflected at the disc surface. $A_{\mathrm{track}}(e_\mathrm{r})$ can be derived by inspection of Fig. 3.9 and by using the rules of areas of disc segments. The expression of $A_{\mathrm{track}}(e_\mathrm{r})$ consists of four parts each supported in an interval of e_r. The four different situations are illustrated in the figures 3.14-3.15. $A_{\mathrm{track}}(e_\mathrm{r})$ is dependent on the absolute value of e_r due to the symmetry of the track and laser spot.
By inspection of these figures the expressions of A_{track} can be found for the four cases:
if $\mathrm{abs}(e_\mathrm{r}) \leq r - \frac{b_\mathrm{s}}{2}$

$$A_{\mathrm{track}}(e_\mathrm{r}) = \pi \cdot r^2 - \frac{r^2}{2} \cdot (\theta_{\mathrm{a},1} - \sin\theta_{\mathrm{a},1}) - \frac{r^2}{2} \cdot (\theta_{\mathrm{b},1} - \sin\theta_{\mathrm{b},1}), \quad (3.44)$$

where

$$\theta_{\mathrm{a},1} = 2 \cdot \arccos \frac{\mathrm{abs}(e_\mathrm{r}) + \frac{b_\mathrm{s}}{2}}{r}, \quad (3.45)$$

$$\theta_{\mathrm{b},1} = 2 \cdot \arccos \frac{\mathrm{abs}(e_\mathrm{r} - \frac{b_\mathrm{s}}{2})}{r}, \quad (3.46)$$



(a) $\text{abs}(e_\text{r}) \leq r - \frac{b_\text{s}}{2}$. (b) $r - \frac{b_\text{s}}{2} < \text{abs}(e_\text{r}) \leq \frac{b_\text{s}}{2}$

Figure 3.14: Illustration of the laser spot placement over the track, the gray area is $A_\text{track}(e_\text{r})$.

if $r - \frac{b_\text{s}}{2} < \text{abs}(e_\text{r}) \leq \frac{b_\text{s}}{2}$

$$A_\text{track}(e_\text{r}) = \pi \cdot r^2 - \frac{r^2}{2} \cdot (\theta_2 - \sin\theta_2), \tag{3.47}$$

where

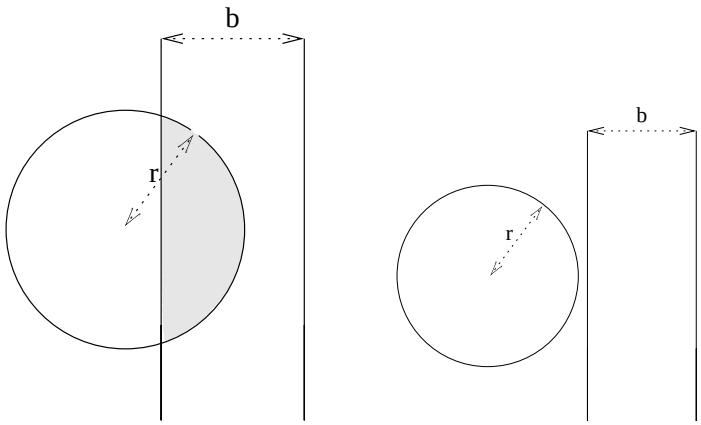
$$\theta_2 = 2 \cdot \arccos \frac{\text{abs}(e_\text{r}) - \frac{b_\text{s}}{2}}{r}, \tag{3.48}$$

if $\frac{b_\text{s}}{2} < \text{abs}(e_\text{r}) \leq r + \frac{b_\text{s}}{2}$

$$A_\text{track}(e_\text{r}) = \frac{r^2}{2} \cdot (\theta_3 - \sin\theta_3), \tag{3.49}$$

where

$$\theta_3 = 2 \cdot \arccos \frac{\text{abs}(e_\text{r}) - \frac{b_\text{s}}{2}}{r}, \tag{3.50}$$



(a) $\frac{b_s}{2} < \text{abs}(e_r) \leq r + \frac{b_s}{2}$. (b) $r + \frac{b_s}{2} \leq \text{abs}(e_r)$.

Figure 3.15: Illustration of the laser spot placement over the track, the gray area is $A_{\text{track}}(e_r)$.

if $r + \frac{b_s}{2} \leq \text{abs}(e_r)$

$$A_{\text{track}} = 0. \tag{3.51}$$

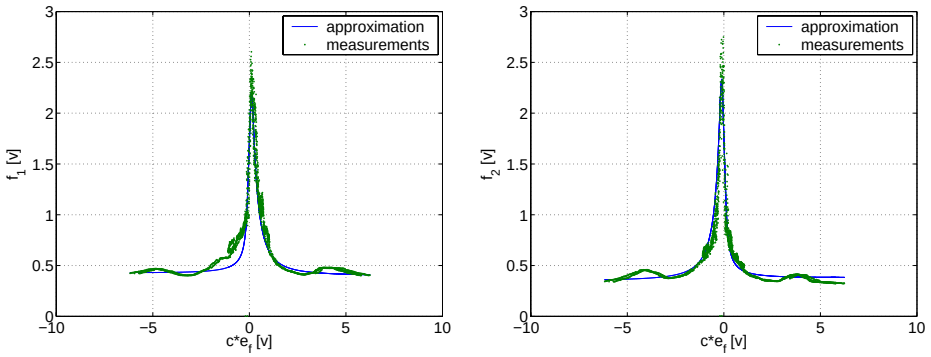
The expression of $h_1(e_r)$ is found by substituting the expressions in (3.44-3.51) into (3.43).

3.4.3 Measurements and parameter identification

These measurements are performed using the test setup described in Section 3.1.
The measurements of e_f are performed by slowly varying the focus position which is done by applying a slowly varying saw tooth signal as control signal, u_f, while the disc does not spin around. This signal is sampled with a frequency at 5 kHz. The measurements of the radial distance's influence are done with the disc spinning around, and with an active focus controller but without radial tracking control. This will result in a harmonic movement of the pick up relative to the disc surface due to the eccentricity of the disc. This signal is sampled with a frequency at 35 kHz. Only a part of this signal is used for parameter identification, see [Odgaard et al., 2003c]. It is a part where the pick-up crosses the track slowly. This means that the derivative of the sine function is zero and the function value is close to zero. The parameters in this sine function are identified along with the model parameters in the parameter identification. The first measurements are used for the identification of parameters in the $h_i(e_f)$ functions and the second set of

measurements are used for $g_i(e_\mathrm{r})$. Before doing the identification the $h_i(e_\mathrm{f})$ and $g_i(e_\mathrm{r})$ are multiplied with a parameter due to the amplification in the detector and an offset is added to this also due to the detectors.

The parameter identifications were done by using the MATLAB™function `fminsearch` to minimise the squared error of the difference between the measurements and model output, by changing the parameters. The initial parameters were found partly in [Stan, 1998] and partly by trial and error. The model obtained with the identified parameters is compared with the measurements in Figs. (3.16-3.19) as functions of either $c \cdot e_\mathrm{f}$ or e_r.



(a) $f_1(c \cdot e_\mathrm{f}, 0)$ (b) $f_2(c \cdot e_\mathrm{f}, 0)$

Figure 3.16: Measurements of $f_1(c \cdot e_\mathrm{f}, 0)$ and $f_2(c \cdot e_\mathrm{f}, 0)$ compared with the simulated $f_1(c \cdot e_\mathrm{f}, 0)$ and $f_2(c \cdot e_\mathrm{f}, 0)$. These two functions model the two focus detector signals dependency of the focus distance.

The Figs. (3.16-3.19) show how the models in general have almost the same response to the given inputs as the optical system. But due to some of the simplifications in the model/assumptions, the models do not describe all phenomena in the optical detector system. In Figs. 3.16(a), 3.16(b), 3.18(a) and 3.18(b) there are some bumps for large displacement, due to light from the other detector beams, which are not modelled. However, these differences are not so important, concerning the purpose of this model, which is in the control area. It can be corrected together with another model problem which emerges in the further use of the optical model. This model is non differentiable at some points, even though the real optical system do not show such a behaviour. This is due to piecewise structure of the model. In [Odgaard et al., 2004a] and [Odgaard et al., 2003a], these problems are handled by approximating the $h_i(\cdot)$ and $g_i(\cdot)$ by cubic splines.

All the identified parameters can be seen in Appendix B.

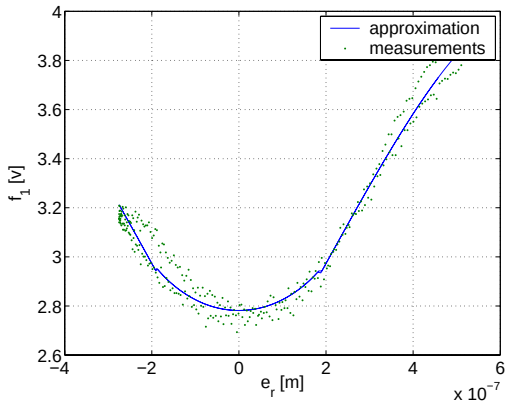


Figure 3.17: The focus detector as a function of the radial distance, $f_1(0, e_\mathrm{r})$, showing the measured and simulated values.

3.5 Surface fault model

The purpose of the surface fault model and its related fault parameters are to use them for fault detection and classification. In the following a realistic and simple model is found. This model is as the optical model a static model, so the time variation of the variables has been disregarded,

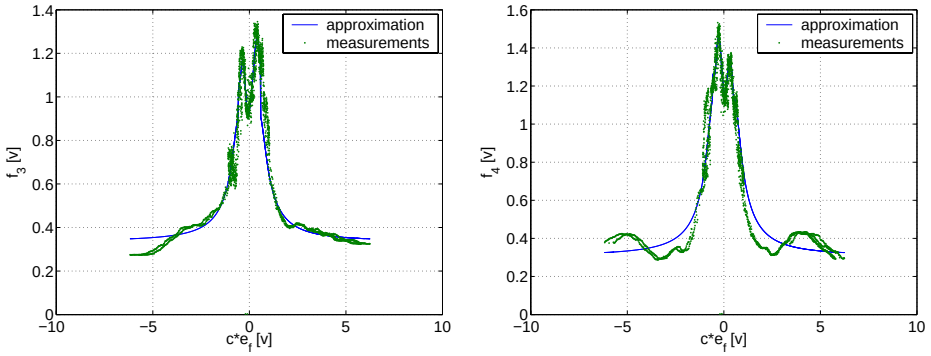
$$\mathbf{s}_\mathrm{m} = \begin{bmatrix} D_{1_m} \\ D_{2_m} \\ S_{1_m} \\ S_{2_m} \end{bmatrix} = \mathbf{f}_\mathrm{e}(e_\mathrm{f}, e_\mathrm{r}, \alpha_\mathrm{f}, \alpha_\mathrm{r}), \tag{3.52}$$

$\mathbf{s}_\mathrm{m}$ is a vector of the measured photo detector signals. This model is related to the optical model (3.17) by the following

$$\mathbf{f}_\mathrm{e}(e_\mathrm{f}, e_\mathrm{r}, 0, 0) = \mathbf{f}(e_\mathrm{f}, e_\mathrm{r}). \tag{3.53}$$

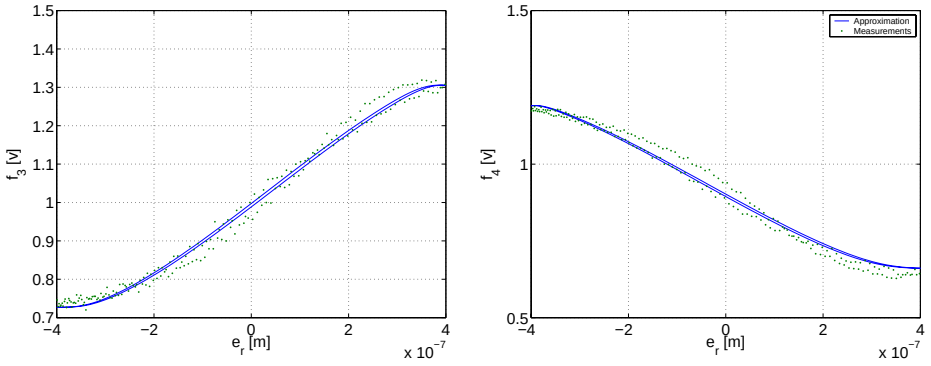
It is seen by inspection that a fault decreases $D_1 + D_2$ and $S_1 + S_2$. An example of $D_1 + D_2$ sequence of a scratchy disc, Fig. 3.20. From this figure is it clear that a fault decreases $D_1 + D_2$.
In [Odgaard et al., 2004a] a disturbance set is defined.

3.1 Definition (The disturbance set) *The disturbance set $\mathcal{D} \in \mathcal{R}^4$ is defined as the set in which any sample $\mathbf{s}_m$ in $\mathcal{R}^4$ of the detector signals, will be if only disturbances occur.*



(a) $f_3(c \cdot e_\mathrm{f}, 0)$. (b) $f_4(c \cdot e_\mathrm{f}, 0)$.

Figure 3.18: The radial detector as functions of the focus distance, $f_3(c \cdot e_\mathrm{f}, 0)$ and $f_4(c \cdot e_\mathrm{f}, 0)$. The figures compare the measured and simulated functions $f_3(c \cdot e_\mathrm{f}, 0)$ and $f_4(c \cdot e_\mathrm{f}, 0)$.



(a) $f_3(0, e_\mathrm{r})$. (b) $f_4(0, e_\mathrm{r})$.

Figure 3.19: An illustration of the comparison of the measured and simulated radial detectors dependence of the radial distances, $f_3(0, e_\mathrm{r})$ and $f_4(0, e_\mathrm{r})$.

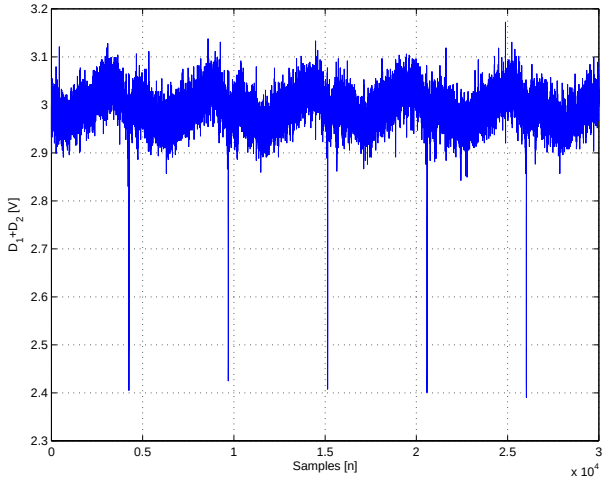


Figure 3.20: A $D_1 + D_2$ sequence of a scratchy disc, notice the spikes which are the scratches.

To illustrate this disturbance set, a 2-dimensional version, where the two focus detectors D_1 and D_2 are plotted in Fig. 3.21. The detector would be in the normal region of $\mathcal{D}$, if the focus distance is kept inside the normal operation region, the linear part of the optical model. A fault can be viewed as a movement of the detector signals outside this disturbance set, $\mathcal{D}$. In order to derive a model of the surface faults it is important to determine the structure behind these movements outside $\mathcal{D}$.
In [Odgaard et al., 2004a] two models are suggested. The first model is based on the assumption that the fault is orthogonal to $\mathcal{D}$, at the point $((e_\mathrm{f}, e_\mathrm{r}), \mathbf{f}(e_\mathrm{f}, e_\mathrm{r}))$. This orthogonal model is illustrated in Fig. 3.22.
The parameters in this fault model are two orthogonal distances: r_f and r_r, defined as:

$$r_\mathrm{f} = \left\| \begin{matrix} D_1 - \hat{D}_1 \\ D_2 - \hat{D}_2 \end{matrix} \right\|, \; r_\mathrm{r} = \left\| \begin{matrix} S_1 - \hat{S}_1 \\ S_2 - \hat{S}_2 \end{matrix} \right\|. \tag{3.54}$$

These two parameters are also residuals, since they are equal to zero when no faults occur and increase as the fault increases.
However, in cases of focus or radial distances unequal to 0 during the fault, the orthogonal model will not be a good description of the fault, since experimental work in the 2-dimensional case, in [Vidal et al., 2001a], indicates what all detector signals tend towards the origin during a fault. This could indicate that it would be better to model the fault with a scale, β, multiplied with $\mathbf{s}$, as the first parameter and γ which is the model error as the second parameter:

$$\mathbf{s}_\mathrm{m} \approx \beta \cdot \mathbf{s}, \;\; \gamma = ||\mathbf{s}_\mathrm{m} - \beta \cdot \mathbf{s}||, \tag{3.55}$$

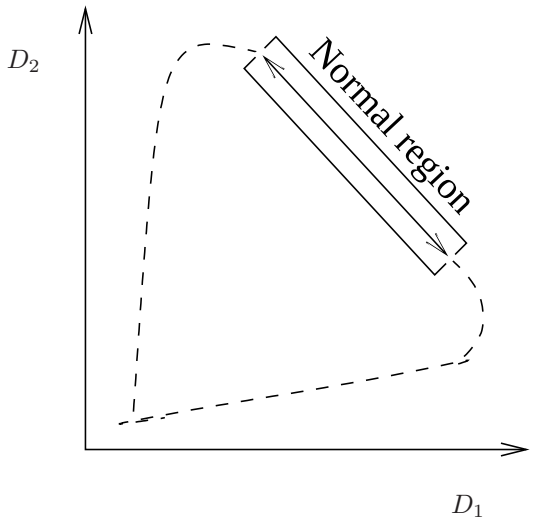


Figure 3.21: A 2-dimensional illustration of the disturbance set. D_2 as a function of D_1 for small focus distances, inside the normal operation region. The dashed line illustrates detector signal path outside the normal region.

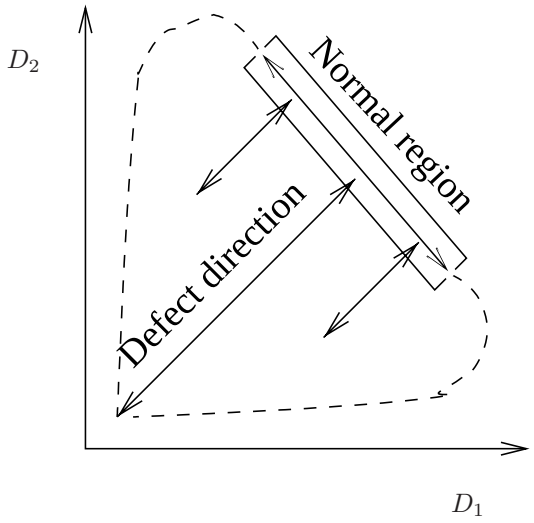


Figure 3.22: A 2-dimensional illustration of an orthogonal fault model. The fault moves the detector signals outside the normal operation region. The dashed line illustrates detector signal path outside the normal region.

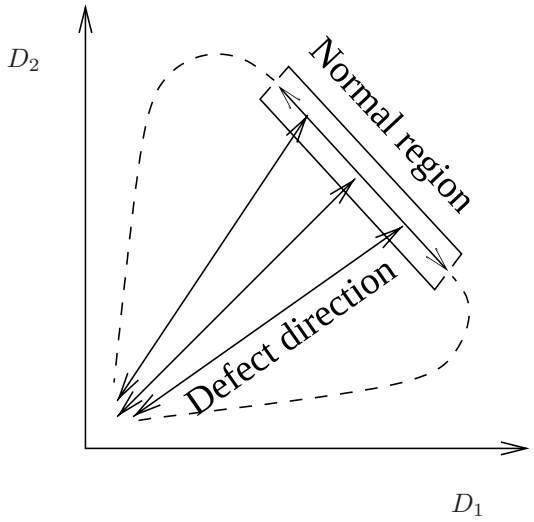


Figure 3.23: A 2-dimensional illustration of a scaling fault model. The fault moves the detector signals outside the normal operation region. The dashed line illustrates detector signal path outside the normal region.

β is not a residual since it is 1 if no fault occurs and decreases as the fault develops. Instead a residual, α, can be found based on β by:

$$\alpha = 1 - \beta. \tag{3.56}$$

This scaling model is illustrated in Fig. 3.23. Since only one fault variable is used, this fault model does not guarantee that $\mathbf{s}_\mathrm{m}$ can be reconstructed based on $e_\mathrm{f}, e_\mathrm{r}, \beta$. As a consequence other model structures were tested. These experiments showed that a well suited model structure is to use one scaling for the focus detectors and another scaling for the radial detectors.

$$\mathbf{s}_\mathrm{m} = \begin{bmatrix} \beta_\mathrm{f} \cdot \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \beta_\mathrm{r} \cdot \mathbf{I} \end{bmatrix} \cdot \mathbf{s}, \tag{3.57}$$

and the residuals based on β_f and β_r can be constructed as

$$\begin{bmatrix} \alpha_\mathrm{f} \\ \alpha_\mathrm{r} \end{bmatrix} = \begin{bmatrix} 1 - \beta_\mathrm{f} \\ 1 - \beta_\mathrm{f} \end{bmatrix}. \tag{3.58}$$

These two residuals can be used for fault detection, with a clear improving potential, see Section 4.4, [Odgaard et al., 2003b] and [Odgaard et al., 2003a].
These fault parameters parameterise the main part of the surface faults. Unfortunately the surface faults also contribute with energy in the focus and radial distances, or in

other words surface faults introduce additional faulty focus and radial distance signal components. If these faulty focus and radial distance components are denoted: $\check{e}_{\mathrm{f}}$ and $\check{e}_{\mathrm{r}}$, a fault model can be defined as

$$\mathbf{s}_{\mathrm{m}} = \begin{bmatrix} \beta_{\mathrm{f}} \cdot \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \beta_{\mathrm{r}} \cdot \mathbf{I} \end{bmatrix} \cdot \mathbf{f}\left(e_{\mathrm{f}} + \check{e}_{\mathrm{f}}, e_{\mathrm{r}} + \check{e}_{\mathrm{r}}\right). \tag{3.59}$$

Notice that β_{f} and β_{r} are very useful for fault detection etc, but it is $\check{e}_{\mathrm{f}}$ and $\check{e}_{\mathrm{r}}$ which cause the problem for the controllers during a surface fault.

3.6 Summary

This chapter was devoted to the modelling of the focus and radial servos in a CD-player, with a main focus on the optical model and a model of surface faults. The model is developed to model the used experimental test setup. This experimental test setup is described in Section 3.1. The general model structure was described in Section 3.2. In Section 3.3 a simple dynamical model of the electro-magnetic-mechanical system was summarised. An often missed part of the focus and radial servos, the optical system was modelled with cross-couplings in Section 3.4. In Section 3.5 a model structure of the surface faults was presented.

Chapter 4

Fault Residuals Based on Estimations of the Optical Model

In this chapter the inverse map of the combined optical and surface fault model is solved. To understand the importance of this inverse map, one should notice the following remarks on the modelling done in the previous chapter. In Chapter 3 the focus and radial servos were modelled. The dynamic model of the electro-magnetic-mechanical system was first derived. Normally the detector signals are used directly to compute the feedback signals and the residuals. These computations are based on simple linear models of the optical system in the OPU, and without any models of the surface faults. The handling of these surface faults is the main topic in this work. A static non-linear model of the optical system was derived in Chapter 3 together with a static model of the surface faults. These two models are defined by mappings from the focus and radial distances and the two fault parameters to the four detector signals. The inverse map of this mapping gives focus and radial distances together with some fault parameters. The fault parameters defined in Section 3.5 are decoupled from the cross-coupling from the other servo loop. The sum signals depend on the respective other focus or radial distances due to the optical cross-couplings. As a consequence it is problematic to use these sum signals as residuals, but they are used due to the lack of better alternatives. This problem with the optical cross-couplings can be avoided by solving the inverse map and instead use the fault parameters and their derived residuals. Unfortunately, the non-linear optical model does not have a globally defined solution. Instead an iterative method to compute the inverse map is used, based on Newton-Raphson's method and a Kalman estimator. In Section 4.1, the structure of this inverse problem solver is described. This is followed by the solution of the inverse map based on an iterative method, see Section 4.2. A Kalman estimator is introduced and designed in Section 4.3, in order to estimate the real focus and radial distances during surface faults. In Section 4.3 the method is used on some experimental data. In Section 4.4 the potentials of the new decoupled residuals are

illustrated by a fault detection example.

4.1 The structure of the inverse map solver

In the normal servo system indirect measures of focus and radial distances are used. These indirect measures of the distances are found by taking the normalised difference between the two focus and the two radial detector signals. It is important to distinguish between disturbances, which the controller shall reject and faults which the controller shall not react to. If preconditioning of the measurements are not used it might make the separation of disturbances from faults more difficult. There are two problems with the measures. The first problem is that there are cross-couplings between the focus distance and the radial detector signals and between the radial distance and focus detector signals. The second problem is that the mappings from the distances to the indirect measures are linear only for small distances. Both these problems lead to the fact that it would be better to establish estimates of focus and radial distances for the controlling of the focus and radial distances, by solving the inverse map of the optical model. It is at least as important that this inverse map can be used to compute some residuals with a clear separation property compared with the normally used ones.

By solving the inverse map of the combined optical and fault model, (3.59), one gets: $\beta_{\mathrm{f}}[n]$, $\beta_{\mathrm{r}}[n]$, $e_{\mathrm{f}}[n] + \check{e}_{\mathrm{f}}[n]$ and $e_{\mathrm{r}}[n] + \check{e}_{\mathrm{r}}[n]$. A Kalman estimator can following be used to estimate the physical focus and radial distances: $e_{\mathrm{f}}[n]$ and $e_{\mathrm{r}}[n]$, which result in the dynamical estimates of focus and radial distances: $\hat{e}_{\mathrm{f}}[n]$ and $\hat{e}_{\mathrm{r}}[n]$. In other words the Kalman estimator opens the closed loop, the two remainders are the open loop faults and disturbances. The two remainders contain more than the faulty distance components due to the surface faults, $\check{e}_{\mathrm{f}}[n]$ and $\check{e}_{\mathrm{r}}[n]$, the remainder also contain measurement noises and disturbances as well. The measurement noises and disturbances can be assumed to be stochastic. The surface faults do not vary much from encounter to encounter, see Fig. 4.1 and [Odgaard et al., 2004d].

The used Kalman estimator is a Kalman estimator with an internal reference model. The internal reference model is a model of the eccentricity and skewness of the disc. The Kalman estimates $\hat{e}_{\mathrm{f}}[n]$ and $\hat{e}_{\mathrm{r}}[n]$ can e.g. be used as partial state estimates to a controller, or fed into the optical model, with the purpose of estimating the detector signals, $\hat{\mathbf{s}}[n]$, the difference from these estimates to the measured distances are open loop estimates of faults and disturbances influencing the system. This dynamical estimate of the detector signals can together with $\tilde{e}_{\mathrm{f}}[n]$ and $\tilde{e}_{\mathrm{r}}[n]$ be used for extraction of the fault parameters and residuals.

A method is suggested for extraction and estimation of focus and radial distances and fault parameters/residuals. These signals can be used to improve the performance of the nominal and the feature based controllers for a CD-player. The system is illustrated in Fig. 4.3. The inverse map solver finds static estimates of optical distances from sampled detector signals.

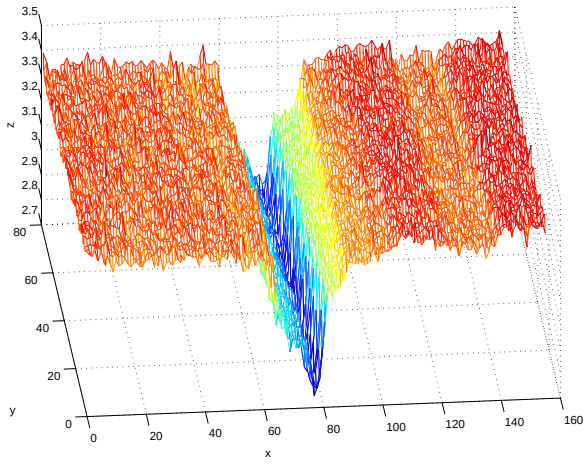


Figure 4.1: 3-D plot of $\beta_{\mathrm{f}}[n]$ in sequential fault encounters. The duration and depth of the fault increases. The x-axis is samples of yth encounter, (the y-axis is the encounter number), and the z-axis is $\beta_{\mathrm{f}}[n]$.

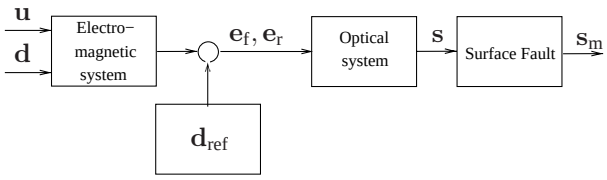


Figure 4.2: The principles of the model of the CD-player. The CD-player consists of four parts. The electro-magnetic-mechanical system and the optical system, the vector of unknown references $\mathbf{d}_{\mathrm{ref}}(t)$ and the surface fault. $\mathbf{u}(t)$ is a vector of the control signal to the electro-magnetic system, $\mathbf{d}(t)$ is a vector of the unknown disturbances to the electro-magnetic-mechanical system, $\mathbf{d}_{\mathrm{ref}}(t)$ is a vector of the unknown reference to the system, $e_{\mathrm{f}}(t)$ is a vector of the focus distance, $e_{\mathrm{r}}(t)$ is a vector of the radial distance, $\mathbf{s}_{\mathrm{m}}(t)$ is a vector of the measured detector signals, and $\mathbf{s}(t)$ is a vector of detector signals without surface faults.

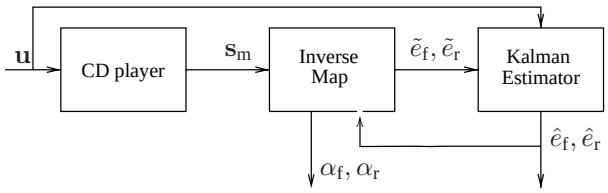


Figure 4.3: The structure of the method described in this chapter. The method consists of two parts, which estimate fault parameters, ($\alpha_{\mathrm{f}}, \alpha_{\mathrm{r}}$), focus distance and radial distance from the detector signals. The parts of the method are: inverse map solver and Kalman estimator. $\mathbf{u}$ is a vector of the control signals to the CD-player, see Fig. 4.2, $\mathbf{s}_{\mathrm{m}}$ is a vector of the measured detector signals, $\tilde{e}_{\mathrm{f}}, \tilde{e}_{\mathrm{r}}$ is the static estimated focus and radial distances, and $\hat{\mathbf{e}}_{\mathrm{f}}, \hat{\mathbf{e}}_{\mathrm{r}}$ is the dynamical estimated focus and radial distances.

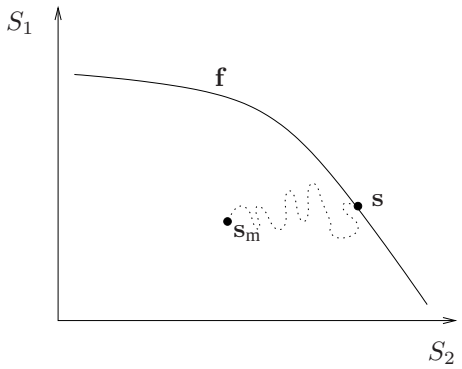


Figure 4.4: Due to a fault the measured detector signal, $\mathbf{s}_{\mathrm{m}}[n]$, lies outside the output set of the optical model, $\mathbf{f}(\cdot)$. This is illustrated with the detector signals S_1 and S_2, where $e_{\mathrm{f}} = 0$ implying that $\mathbf{f}(\cdot)$ is a curve in the S_1-S_2 plane and constant in D_1 and D_2.

The method consists of two parts each depending on each other. The first part is the inverse map of the optical and the fault model.
The four detector signals can be described as in (3.59). $\mathbf{f}_{\mathrm{e}}(\cdot)$ is an extended model combining the optical model and a fault model. $\mathbf{f}(\cdot)$ is the optical model, which maps into a 2-dimensional manifold embedded in a 4-dimensional space.
Combining the model and the four detector signals gives four equations with 2 variables concerning the computation of the two distances. In order to describe the 4-dimensional space two extra variables are used to describe the deviation. These variables are the implicit fault model, parameterised by $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$. This means that the solution to this problem is to compute the inverse map of $\mathbf{f}_{\mathrm{e}}(\cdot)$. The inverse is not globally defined, meaning that the solver only finds a local solution.
In Definition 3.1 in Section 3.5 the disturbance set $\mathcal{D} \in \mathcal{R}^4$ was defined.

The extended model output for two detector signals is illustrated in Fig. 4.4. This figure illustrates a sampling of S_1 and S_2, outside $\mathcal{D}$ due to a fault.

4.2 Solutions of the inverse map

The solution of the inverse map can be found based on the use of the fault model and the optical model. The method and algorithm are described in [Odgaard et al., 2004a] and [Odgaard et al., 2003a]. The method is based on the following facts. Due to low pass filtering nature of the OPU the following limits are relevant: $|e_{\mathrm{f}}[n]-e_{\mathrm{f}}[n+1]|<\mu_{\mathrm{f}}$ and $|e_{\mathrm{r}}[n]-e_{\mathrm{r}}[n+1]|<\mu_{\mathrm{r}}$, where n is any sample, and μ_{f} and μ_{r} are the maximum changes in $e_{\mathrm{f}}[n]$ and $e_{\mathrm{r}}[n]$, see [Odgaard et al., 2004a] and [Odgaard et al., 2003a]. Based on the standard requirements to the CD-player servos, see [Stan, 1998], the maximum deviation of focus and radial positions from sample to sample can be calculated to: $\mu_{\mathrm{f}}=\mu_{\mathrm{r}}\approx 0.014\mu$m.
In the following an approach is described, which locally computes the inverse map and thereby find the right candidate point, and use this to compute the fault parameters and related residuals.

4.2.1 The scaling projection method

The problem is to find the point, $\mathbf{f}(\tilde{\mathbf{e}}[n])$ in $\mathcal{D}$, where $\tilde{\mathbf{e}}[n]=\begin{bmatrix}\tilde{e}_{\mathrm{f}}[n] & \tilde{e}_{\mathrm{r}}[n]\end{bmatrix}^T$. If the model output is substituted into (3.57) the following is achieved

$$\mathbf{s}_{\mathrm{m}}[n]=\mathcal{B}[n]\cdot\mathbf{f}(\tilde{\mathbf{e}}[n]), \tag{4.1}$$

where

$$\mathcal{B}[n]=\begin{bmatrix}\beta_{\mathrm{f}}[n]\cdot\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \beta_{\mathrm{r}}[n]\cdot\mathbf{I}\end{bmatrix}. \tag{4.2}$$

For computational reasons, a new matrix $\mathcal{B}^*[n]$ is defined,

$$\mathcal{B}^*[n]=\begin{bmatrix}\beta_{\mathrm{f}}^{-1}[n]\cdot\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \beta_{\mathrm{r}}^{-1}[n]\cdot\mathbf{I}\end{bmatrix}. \tag{4.3}$$

Unfortunately, due to the shape of $\mathcal{D}$ in $\mathcal{R}^4$, there are a number of points in $\mathcal{D}$ for which (4.1) will be true. However, due to the system's low-pass-filtering effect $\mathbf{e}[n]$ is expected to be near $\mathbf{e}[n-1]$, this means that if an iterative algorithm is used to find $\tilde{\mathbf{e}}[n]$, $\tilde{\mathbf{e}}[n-1]$ can be used as a starting value of the algorithm. Experiments have shown that it improves results if the dynamically estimated distances are used as this start value. I.e: $\mathbf{e}_0[n]=\begin{bmatrix}\hat{e}_{\mathrm{f}}[n] & \hat{e}_{\mathrm{r}}[n]\end{bmatrix}^T$. Using these low-pass-filtered-estimated distances as initial values for the algorithm, noises etc. are removed from the initial value. This results in

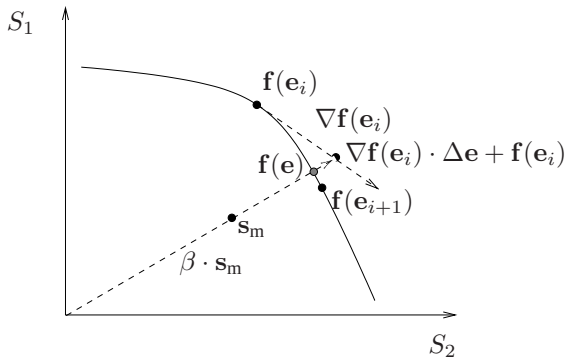


Figure 4.5: Illustration of the principles of the scaling projection method. The Illustration shows an example in $\mathcal{R}^2$. $\mathbf{f}(\mathbf{e}_i)$ is the starting point for the iteration, $\nabla\mathbf{f}(\mathbf{e}_i)$ is the gradient at the starting point, $\mathbf{s}_\mathrm{m}$, $\beta\cdot\mathbf{s}$ is vector through the origin and $\mathbf{s}_\mathrm{m}$, $\nabla\mathbf{f}(\mathbf{e}_i)\cdot\Delta\mathbf{e}+\mathbf{f}(\mathbf{e}_i)$ are the intersection between the tangent plane through $\mathbf{f}(\mathbf{e}_i)$ and the vector through $\mathbf{s}_\mathrm{m}$. This can be used to find $\Delta\mathbf{e}$ and $\mathbf{e}_{i+1}=\mathbf{e}_i+\Delta\mathbf{e}$, so $\mathbf{f}(\mathbf{e}_{i+1})$ is the new function value. The algorithm stops when $\mathrm{norm}(\mathbf{f}(\mathbf{e}_{i+1})-\beta\cdot\mathbf{s})<\epsilon$, where ϵ is the stopping parameter. $\mathbf{f}(\mathbf{e})$ is the vector of detector signals if no fault would had occurred.

the fact that the algorithm requires in general fewer iterations before the estimation error becomes as low as required.

In the following an algorithm, based on Newton's method, is described. One iteration, of the algorithm applied to the problem illustrated in Fig. 4.4, which is a problem in $\mathcal{R}^2$, is illustrated in Fig. 4.5. This figure illustrates a situation of S_1 and S_2 where $e_\mathrm{f}=0$.

First the crossing point between the tangent plane of $\mathcal{D}$ at the initial point and the line which goes through origin and the measurement, is found.

$$\mathbf{f}(\tilde{\mathbf{e}}_i[n])+\nabla\mathbf{f}(\tilde{\mathbf{e}}_i[n])\cdot\Delta\tilde{\mathbf{e}}_{i+1}[n]=\mathcal{B}^*[n]\cdot\mathbf{s}_\mathrm{m}[n]\Rightarrow \tag{4.4}$$

$$\nabla\mathbf{f}(\tilde{\mathbf{e}}_i[n])\cdot\Delta\tilde{\mathbf{e}}_{i+1}[n]-\mathcal{B}^*[n]\cdot\mathbf{s}_\mathrm{m}[n]=-\mathbf{f}(\tilde{\mathbf{e}}_i[n]) \tag{4.5}$$

denote:

$$\mathbf{S}_\mathrm{m}[n]=\begin{bmatrix}D_1[n] & 0\\ D_2[n] & 0\\ 0 & S_1[n]\\ 0 & S_2[n]\end{bmatrix}, \tag{4.6}$$

$$\begin{bmatrix}\nabla\mathbf{f}(\tilde{\mathbf{e}}_i[n]) & -\mathbf{S}_\mathrm{m}[n]\end{bmatrix}\begin{bmatrix}\Delta\tilde{\mathbf{e}}_{i+1}[n]\\ \beta_{\mathrm{f},i}^{-1}[n]\\ \beta_{\mathrm{r},i}^{-1}[n]\end{bmatrix}=-\mathbf{f}(\mathbf{e}_i[n]) \tag{4.7}$$

The left most matrix in (4.7) is in practice invertible, since it has full rank, except from a few situations of low dimension. These situations can be avoided if the initial point for the iterations is in the nominal range of operations.

$$\begin{bmatrix} \Delta\mathbf{e}_{i+1}[n] \\ \beta_{\mathrm{f},i}^{-1}[n] \\ \beta_{\mathrm{r},i}^{-1}[n] \end{bmatrix} = -\left(\begin{bmatrix} \nabla\mathbf{f}(\tilde{\mathbf{e}}_i[n]) & -\mathbf{S}_{\mathrm{m}}[n] \end{bmatrix}\right)^{-1} \cdot \mathbf{f}(\tilde{\mathbf{e}}_i[n]). \tag{4.8}$$

Using all four parameters to model the received detector signals, the detector signals will be fully parameterised and is possible to achieve a zero model error. The algorithm will converge to the solution if it is in the vicinity of the solution, which it is due to the limited change in focus and radial distances from sample to sample. This means that the suggested algorithm will converge as follows: $\lim_{i\to\infty} \mathbf{e}_i[n] = \mathbf{e}[n], \lim_{i\to\infty} e_{\mathrm{f},i}[n] = e_{\mathrm{f}}[n], \lim_{i\to\infty} e_{\mathrm{r},i}[n] = e_{\mathrm{r}}[n]$.
It is thereby relevant to use the L_2 norm of the difference between estimated and measured detector signals, as the stop criteria for the algorithm:

$$\left\|(\mathbf{f}(\mathbf{e}[n]) - \beta[n] \cdot \mathbf{s}_{\mathrm{m}}[n])\right\|_2 < \epsilon. \tag{4.9}$$

If the stop criteria is fulfilled, the algorithm stops, and if not fulfilled, the computed distances are used as initial guess for the next iteration. When the algorithm is stopped the residuals are computed.

From fault parameters to residuals In fault detection residuals are used to detect faults (surface faults). In this terminology a residual means more than just a difference between two signals, e.g. the difference between measured and estimated system outputs. In [Chen and Patton, 1991] a residual is defined to be zero for the fault-free case and nonzero for the fault free case. In this case the fault parameters $\beta_{\mathrm{f}}[n]$ and $\beta_{\mathrm{r}}[n]$ are equal to 1 in the fault free case and decreases towards 0 in case of a fault. The fault parameters can be converted to residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$

$$\begin{bmatrix} \alpha_{\mathrm{f}}[n] \\ \alpha_{\mathrm{r}}[n] \end{bmatrix} = \begin{bmatrix} 1 - \beta_{\mathrm{f}}[n] \\ 1 - \beta_{\mathrm{r}}[n] \end{bmatrix}. \tag{4.10}$$

4.2.1.1 The algorithm

1. Find the gradient, $\nabla\mathbf{f}(\tilde{\mathbf{e}}_i[n])$ to the point $(\tilde{\mathbf{e}}_i[n], \mathbf{f}(\tilde{\mathbf{e}}_i[n]))$.
2. Compute:
$$\begin{bmatrix} \tilde{\mathbf{e}}_{i+1}[n] \\ \beta_{\mathrm{f},i}[n] \\ \beta_{\mathrm{r},i}[n] \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{e}}_i[n] \\ 0 \\ 0 \end{bmatrix} - \left(\begin{bmatrix} \nabla\mathbf{f}(\tilde{\mathbf{e}}_i[n]) & -\mathbf{S}_{\mathrm{m}}[n] \end{bmatrix}\right)^{-1} \cdot \mathbf{f}(\tilde{\mathbf{e}}_i[n]).$$
3. Compute: $\gamma[n] = \left\|\mathcal{B}_i[n]\mathbf{s}_{\mathrm{m}}[n] - \mathbf{f}(\mathbf{e}_{i+1}[n])\right\|_2$

4. Jump to step 1 if: $\epsilon < \gamma[n]$, else stop.

5. Compute the residuals: $\begin{bmatrix} \alpha_{\mathrm{f}}[n] \\ \alpha_{\mathrm{r}}[n] \end{bmatrix} = \begin{bmatrix} 1 - \beta_{\mathrm{f},i}[n] \\ 1 - \beta_{\mathrm{f},i}[n] \end{bmatrix}$.

Regarding the convergence properties of the algorithm the optical model is approximated by four products of second and third order polynomiums. It is a known fact that the Newton-Raphson's method converges then it solves scalar second or third order equations. However, since the model is not scalar, it is not known if convergence can be guaranteed, but assumed due to Newton-Raphson performance on polynomiums. In practise the algorithm has shown very good convergence, when tested. It has converged within 3 iterations, with $\epsilon = 10^{-12}$.

4.2.2 Simulation of the inverse map solution

In this section the ability of the algorithm to solve the described inverse map is tested by a number of different simulations. The input signal to this simulation is chosen as two harmonic signals with a small difference in the frequency so that the two input signals are not fully correlated, see [Odgaard et al., 2004a] and [Odgaard et al., 2003a]. The frequency and amplitude of the these sine signals are chosen in a way such that the maximum difference value between two samples is at least 0.014μm, see the beginning of Section 4.2. Starting with simulating the output of the optical model without any faults, see Fig. 4.6. The simulated faults are constructed based on the model in (3.57),

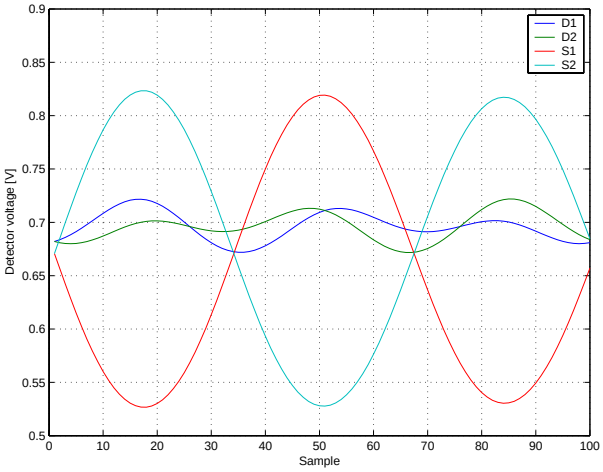


Figure 4.6: Simulated optical detector signals without faults used for simulation of the inverse map solver.

where $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$ model the scratches on the surface, and the signals are illustrated

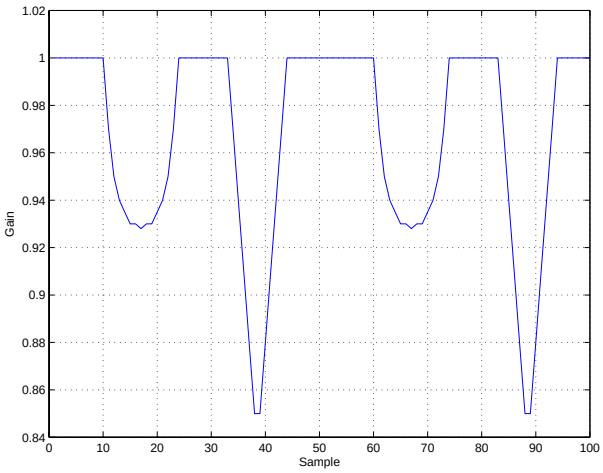


Figure 4.7: The focus and radial residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$ time series, which are used as the surface faults in this simulation.

in Fig. 4.7, in this simulation $\alpha_{\mathrm{f}}[n] = \alpha_{\mathrm{r}}[n]$. Using the fault model and the fault signal in Fig. 4.7, the simulation series of samples with surface faults are computed and illustrated in Fig. 4.8. Before using the algorithm on these test data, independent random noises are added to the four detector signals, to simulate disturbances and measurement noises. The next step in the simulation is to apply the signal from Fig. 4.8 to the algorithm. In this simulation the detector signals shown in Fig. 4.8 are fed to the algorithm. The results are shown in Fig. 4.9. In Fig. 4.9 $\mathbf{f}(\tilde{\mathbf{e}}[n])$ is illustrated. From this plots it is clear that the scaling projection method is well suited for solving this inverse map with the given fault model.

4.3 Kalman estimator

In Fig. 4.9 the measurement noises, faulty focus and radial distances components are remaining after the solution of the inverse map. The computed focus and radial distances are when mapped through the optical model before its appearance in Fig. 4.9. In order to remove these measurement noises and faulty focus and radial distance components from the computed focus and radial distances, a Kalman estimator is designed to estimate the focus and radial distances as a response of the control signals. In addition this Kalman estimator can be used to give an estimate of the faults and disturbances with the system response to these.
The reference signals to focus and radial loops are unknown. The nature of these ref-

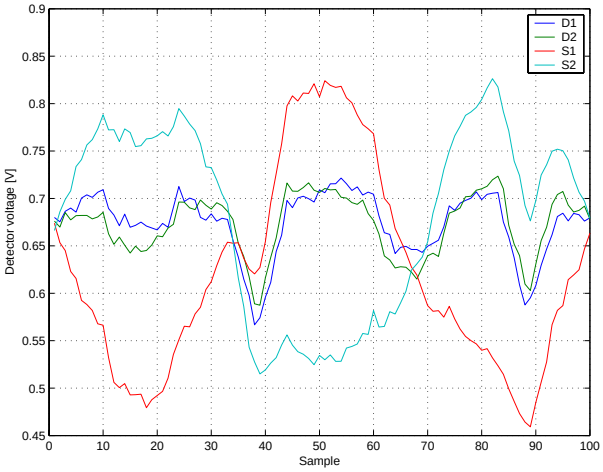


Figure 4.8: Simulation of the four detector signals, $D_1[n]$, $D_2[n]$, $S_1[n]$ and $S_2[n]$, with surface faults and measurement noises.

erences is eccentricity and skewness of disc etc. However, it is known that the first and most dominant harmonic of the unknown reference is the angular velocity of the disc. The interval of the angular velocity can be retrieved from [Stan, 1998] to be $4 - 9$Hz. One way to model the reference is by a bandpass filter with the bandpass region between 4 and 9 Hz. The model will be of the structure:

$$\dot{\eta}_{\text{ref}}(t) = \mathbf{A}_{\text{ref}} \cdot \eta_{\text{ref}}(t) + \mathbf{E}_{\text{ref}} \cdot q_{\text{ref}}(t), \tag{4.11}$$

$$d_{\text{ref}}(t) = \mathbf{C}_{\text{ref}} \cdot \eta_{\text{ref}}(t), \tag{4.12}$$

where: $\eta_{\text{ref}}(t)$ is the state vector of the internal reference model, $d_{\text{ref}}(t)$ is the reference, $q_{\text{ref}}(t)$ is the input to internal reference model which is a white noise signal.

The focus and radial electro-magnetic-mechanical models given in (3.11 -3.12) are transform to a states-pace representation. It is following possible to merge the internal reference model with focus and radial models.

$$\dot{\eta}_{\text{CD}}(t) = \mathbf{A}_{\text{CD}} \cdot \eta_{\text{CD}}(t) + \mathbf{B}_{\text{CD}} \cdot \mathbf{u}_{\text{CD}}(t) + \mathbf{E} \cdot \mathbf{q}_{\text{ref}}(t), \tag{4.13}$$

$$\begin{bmatrix} e_{\text{f}}(t) \\ e_{\text{r}}(t) \end{bmatrix} = \mathbf{C}_{\text{CD}} \cdot \eta_{\text{CD}}(t), \tag{4.14}$$

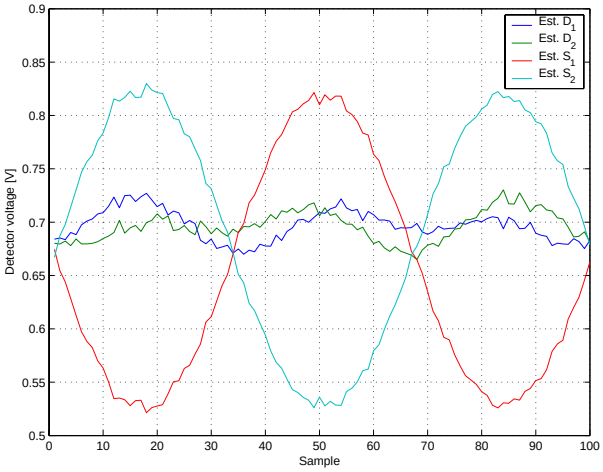


Figure 4.9: Illustration of the estimation of the four detector signals, $D_1[n]$, $D_2[n]$, $S_1[n]$ and $S_2[n]$ by the use of the scaling projection method.

where

$$\mathbf{A}_{\text{CD}} = \begin{bmatrix} \mathbf{A}_{\text{f}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{\text{r,f}} & \mathbf{0} & \mathbf{0}, \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{\text{r}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{\text{r,r}} \end{bmatrix} \in \mathcal{R}^{8\times 8}, \tag{4.15}$$

$$\mathbf{B}_{\text{CD}} = \begin{bmatrix} \mathbf{B}_{\text{f}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{\text{r}}, \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathcal{R}^{8\times 2}, \tag{4.16}$$

$$\mathbf{C}_{\text{CD}} = \begin{bmatrix} \mathbf{C}_{\text{f}} & \mathbf{C}_{\text{r,f}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{\text{r}} & \mathbf{C}_{\text{r,r}} \end{bmatrix} \in \mathcal{R}^{2\times 8}, \tag{4.17}$$

$$\mathbf{E} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{E}_{\text{r,f}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \\ \mathbf{0} & \mathbf{E}_{\text{r,r}} \end{bmatrix} \in \mathcal{R}^{8\times 2}, \tag{4.18}$$

where $\mathbf{A}_{\text{f}}, \mathbf{B}_{\text{f}}, \mathbf{C}_{\text{f}}$ are the model matrices in the focus model, and $\mathbf{A}_{\text{r}}, \mathbf{B}_{\text{r}}, \mathbf{C}_{\text{r}}$ are the model matrices in the radial model. $\mathbf{A}_{\text{r,f}}, \mathbf{C}_{\text{r,f}}, \mathbf{E}_{\text{r,f}}$ are the matrices of the focus reference model, and $\mathbf{A}_{\text{r,r}}, \mathbf{C}_{\text{r,r}}, \mathbf{E}_{\text{r,r}}$ are the matrices of the radial reference model.
Before designing the Kalman estimator, the dynamical model of the electro-magnetic-mechanical system is discretised and expanded to include state and measurement noises

as well. This gives the following model:

$$\eta[n+1] = \mathbf{\Phi} \cdot \eta[n] + \mathbf{\Gamma} \cdot \mathbf{u}[n] + \mathbf{\Psi} \cdot \begin{bmatrix} w_1[n] \\ w_2[n] \end{bmatrix}, \tag{4.19}$$

$$\begin{bmatrix} e_{\mathrm{f}}[n] \\ e_{\mathrm{r}}[n] \end{bmatrix} = \mathbf{C}_{\mathrm{CD}} \cdot \eta[n] + \begin{bmatrix} v_1[n] \\ v_2[n] \end{bmatrix}, \tag{4.20}$$

where: $\mathbf{\Phi}, \mathbf{\Gamma}$ are the discrete time version of $\mathbf{A}_{\mathrm{CD}}, \mathbf{B}_{\mathrm{CD}}$. $w_1[n], w_2[n], v_1[n], v_2[n]$ are independent noises. In the design it is assumed that the $w_1[n], w_2[n]$ can be viewed as noises added to the control signal. The variance of $\mathbf{w}[n] = \begin{bmatrix} w_1[n] \\ w_2[n] \end{bmatrix}$ is named $\mathbf{Q}$, the variance of $\mathbf{v}[n] = \begin{bmatrix} v_1[n] \\ v_2[n] \end{bmatrix}$ is named $\mathbf{R}$, and the covariance is assumed to be zero. These variance values are unknown and are consequently used as tuning parameters in the design process of the Kalman estimator. Due to the fact that the CD-player will run for a very long time, it is chosen to design a steady state Kalman estimator, see [Franklin et al., 1998]. I.e. the steady state Kalman gain, $\mathbf{L}_\infty$, and steady state innovation gain, $\mathbf{M}_\infty$, are to be found. The estimator equations are seen in (4.21)-(4.22):

$$\hat{\eta}[n+1] = \mathbf{\Phi} \cdot \hat{\eta}[n] + \mathbf{\Gamma} \cdot \mathbf{u}[n] + \mathbf{L}_\infty \cdot \left(\begin{bmatrix} \tilde{e}_{\mathrm{f}}[n] \\ \tilde{e}_{\mathrm{r}}[n] \end{bmatrix} - \mathbf{C}_{\mathrm{CD}} \cdot \hat{\eta}[n] \right), \tag{4.21}$$

$$\begin{bmatrix} \hat{e}_{\mathrm{f}}[n] \\ \hat{e}_{\mathrm{r}}[n] \end{bmatrix} = \mathbf{C}_{\mathrm{CD}} \cdot (\mathbf{I} - \mathbf{M}_\infty \cdot \mathbf{C}_{\mathrm{CD}}) \cdot \hat{\eta}[n] + \mathbf{C}_{\mathrm{CD}} \cdot \mathbf{M}_\infty \cdot \begin{bmatrix} \tilde{e}_{\mathrm{f}}[n] \\ \tilde{e}_{\mathrm{r}}[n] \end{bmatrix}. \tag{4.22}$$

The two gains are calculated by use of the MATLAB™function *Kalman*, these gains can be seen in Appendix C. This function has the variance matrices as inputs, and they are found during an iterative design process of the Kalman estimator, such that the Kalman estimator do not remove the low frequently disturbances In Fig. 4.10 a bode plot of the Kalman estimators transfer function from $\tilde{e}_{\mathrm{f}}[n]$ to $\hat{e}_{\mathrm{f}}[n]$ is shown. Experimental work indicates that disturbances are not significantly present above 15 krad/s. This is reflected by the high frequent roll off in the Kalman filter Bode plot in Fig. 4.10. The same experimental work indicates that surface faults have frequency content in the range from 3 krad/s. The gain of the observer is only high in the frequency range for which the disturbances are present. However, a part of this range is also in the frequency range of the surface faults. This means that the estimator contains the disturbances and removes a large part of the fault signal from the estimate.

The output of the simulation of inverse map solver in Fig. 4.9 shows the inverse map solver's lacking ability to remove noise etc, from the estimated focus and radial distances. The Kalman estimator is designed to remove those. This is simulated by feeding the Kalman estimator with signals as shown in Fig. 4.10. These statically and dynamically estimated focus distances can be seen in Fig. 4.11 and the radial distances in Fig. 4.12.

From Figs. 4.11 and 4.12 it can be seen that the Kalman estimator removes the disturbances, but introduces a phase shift. The reason for this phase shift, is that the frequency

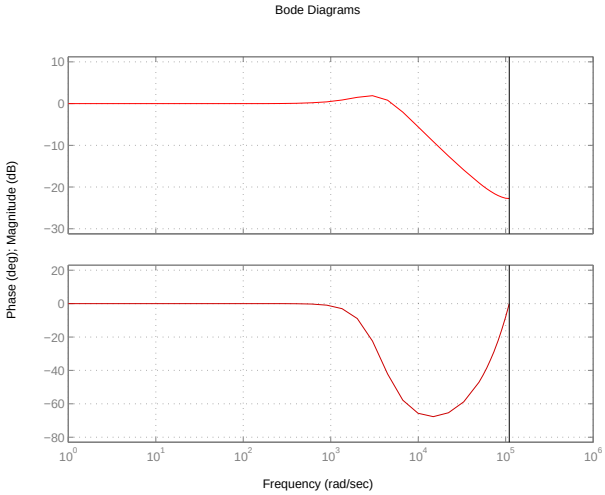


Figure 4.10: Bode plot of the Kalman estimator's transfer function from $\tilde{e}_{\mathrm{f}}[n]$ to $\hat{e}_{\mathrm{f}}[n]$.

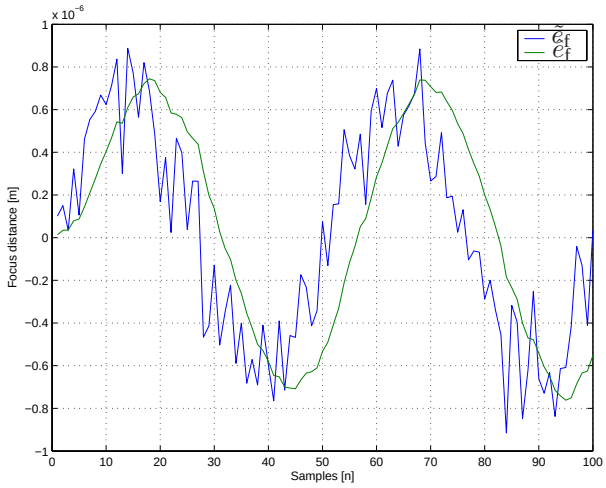


Figure 4.11: The statically and dynamically estimated focus distance signals used in the simulation of the inverse mapping solver.

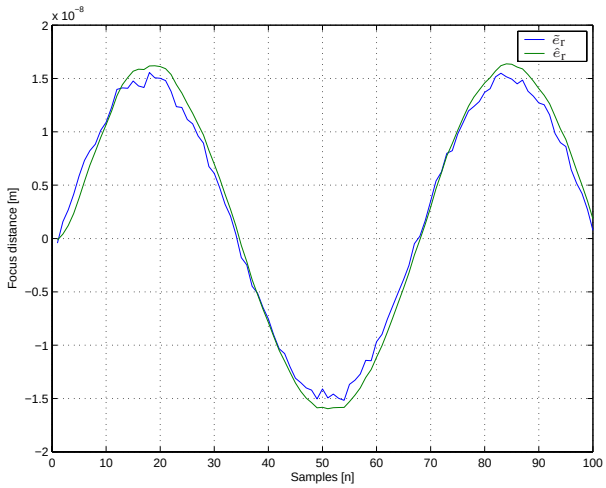


Figure 4.12: The static and dynamic estimated radial distance signals used in the simulation of the inverse mapping solver.

of input signal to the inverse map solver was higher than the assumed maximum frequency in the disturbances, meaning the frequency in the disturbance introduces a phase shift in the Kalman estimator, which can be seen by Fig. 4.11. The frequency in these test signals were chosen so high in order to challenge the the inverse map solver. In practice the disturbance energy in this high frequency would be neglectable, and the main content be in lower frequencies where no phase shift occurs.
The Kalman estimator estimates $\hat{e}_{\mathrm{f}}[n] \approx e_{\mathrm{f}}[n]$ and $\hat{e}_{\mathrm{r}}[n] \approx e_{\mathrm{r}}[n]$, and the difference between the static and dynamic estimates of focus and radial distances is equal the faulty distance component added with some noises. This means that $\tilde{e}_{\mathrm{f}}[n] - \hat{e}_{\mathrm{f}}[n]$ is not equal to $\check{e}_{\mathrm{f}}[n]$, as well as $\tilde{e}_{\mathrm{r}}[n] - \hat{e}_{\mathrm{r}}[n]$ is not equal to $\check{e}_{\mathrm{r}}[n]$, since noise components, $\mathbf{m}_{\mathrm{n}}$, are added to these differences.

4.4 Decoupled residuals and improved time localisation

In Chapter 3 a pair of optical decoupled residuals were derived, and computed in Section 4.2. The question is: are these new residuals more suitable than the normal used residuals for fault localisation and feature extraction. In [Odgaard et al., 2003a] a simple thresholding method is used to compare the new residuals with the standard residuals. The threshold detection method has detected a fault at sample n if $f_{\mathrm{d}}[n] = 1$, and not detected a fault if $f_{\mathrm{d}}[n] = 0$. $f_{\mathrm{d}}[n]$ is defined in (4.23).

$$f_{\mathrm{d}}[n] = \left\| f_{\mathrm{d,f}}[n] \quad f_{\mathrm{d,r}}[n] \right\|_{\infty}, \tag{4.23}$$

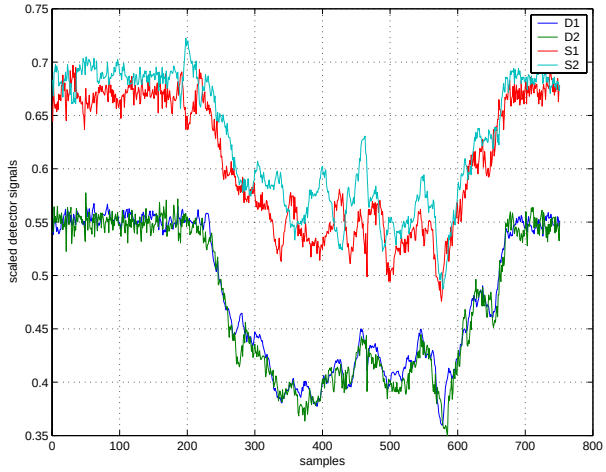


Figure 4.13: Measured detector signals $D_1[n], D_2[n], S_1[n]$ and $S_2[n]$ while passing the scratch.

where

$$f_{\mathrm{d},i}[n] = \begin{cases} 1 & \text{if } r_i[n] > \lambda_i, \\ 0 & \text{if } r_i[n] < \lambda_i, \end{cases} \tag{4.24}$$

$$i \in \{\mathrm{f}, \mathrm{r}\}, \tag{4.25}$$

where $r[n]$ is the residual tested, in the tests two sets of residuals are tested, $\{\alpha_\mathrm{f}[n], \alpha_\mathrm{r}[n]\}$ and $\{S_\mathrm{f}[n], S_\mathrm{r}[n]\}$. λ_i is the thresholds. $\{S_\mathrm{f}[n], S_\mathrm{r}[n]\}$ are the normally used residuals for surface fault detection - they are defined as follows $S_\mathrm{f}[n] = D_1[n] + D_2[n]$ and $S_\mathrm{r}[n] = S_1[n] + S_2[n]$. This pair of residuals are also a representation of the light reflection rate, but without taking cross couplings and non-linearities into account. The new residuals are compared to these coupled residuals by the used of a normal thresholding algorithm. The original scratch can be seen in Fig. 4.13. These results can be seen in Figs. 4.14 and 4.15. Fig. 4.14 illustrates the localisation based on the normal and new focus residual, and Fig. 4.15 illustrates the localisation based on the normal and new radial residual. The original scratch can be seen in Fig. 4.15. The thresholds for the given signals are found such that they detect as much of the fault as possible and without making any false detections on the test data. These thresholds are found by a trial and error method. Starting with Fig. 4.13 which illustrates the scaled detector signals, since it is easy from this figure to do a visual detection of the scratch. The real scratch is the part of the signals where the values are decreased. It lasts approximately from sample 230 to sample 670. The ideal fault detection algorithm will end up with these beginning and end detections of the scratch which, however, is a

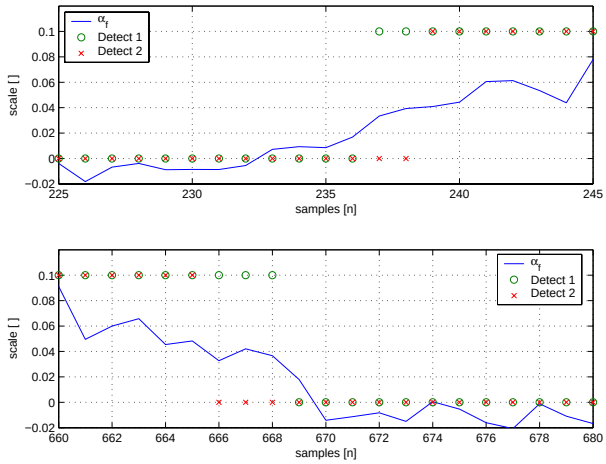


Figure 4.14: $\alpha_{\mathrm{f}}[n]$ plotted together with detection signals based on both $\alpha_{\mathrm{f}}[n]$ (Detect 1) and $S_{\mathrm{f}}[n]$ (Detect 2). The upper plot shows the beginning detection of the scratch, and the lower plot shows the end detection

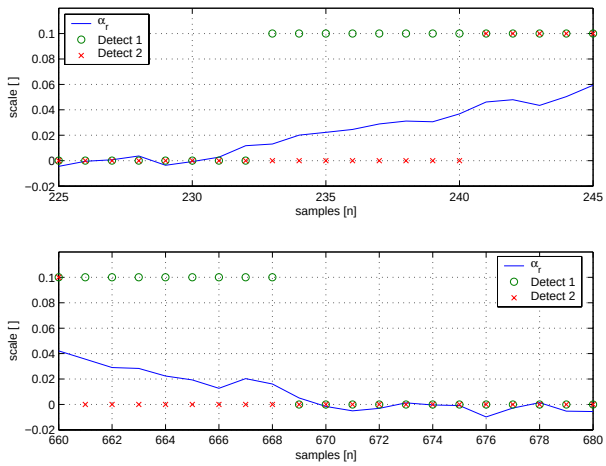


Figure 4.15: $\alpha_{\mathrm{r}}[n]$ plotted together with detection signals based on both $\alpha_{\mathrm{r}}[n]$ (Detect 1) and $S_{\mathrm{r}}[n]$ (Detect 2). The upper plot shows the beginning detection of the scratch, and the lower plot shows the end detection.

hard requirement to fulfill. Figs. 4.14 and 4.15 illustrates the beginning and end of the scratch seen in Fig. 4.13.
From Figs. 4.14 and 4.15 it can be seen that $S_{\mathrm{f}}[n]$, $S_{\mathrm{r}}[n]$, $\alpha_{\mathrm{r}}[n]$ and $\alpha_{\mathrm{r}}[n]$ are good fault parameters/ residuals for detection of the fault, since the size of the variations in the residuals during the scratch are significantly larger than the variations due to noises outside the scratch. Based on these residuals a threshold detection is implemented, by using a threshold which gives the earliest detection of the scratch without giving false detections, (the thresholds were found by lowering these thresholds until false detections were present). The detections based on the four residuals are: the αs : $n = [233 - 668]$, the Ss : $n = [239 - 665]$. From these it can be seen that fault detection based on the new decoupled residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{f}}[n]$ give a more clear detection than if the two other residuals were used, since the background noise level is lower. The new residuals have an improvement of 6 samples in the beginning and 3 samples in the end. This improvement seems less significant compared to duration of the scratch which is 441 samples, but many practical experiments with fault tolerant control in CD-players have shown that an improvement of a few samples of the fault detection means a lot in the achieved controller performances. One should also note that the new method for detection is very close to the fault localisation by the visual inspection. If these two methods are compared one can see that the extended threshold method has an error of 3 samples in the beginning and only 2 sample in the end, on the other hand the fault detection based on the normal residuals had an error of 9 samples in the beginning and 5 in the end.
In addition $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$ are by construction decoupled from the respective radial and focus distances, see [Odgaard et al., 2004a]. As it has been shown earlier the sum signals have a non-linear coupling. In this example the thresholds used for the sum signals were not chosen in the way that the detection is robust towards the optical cross-couplings, meaning that if the detection should be robust towards the cross-couplings the threshold would need to be increased. In turn this would imply that the beginning of the fault would be detected later and the end earlier.
This example makes it clear that it worthwhile to use these new residuals in the time localisaton and feature extraction of the faults in the following chapters, this time localisation and feature extraction will be the topic, starting with some required theory in the field of time/frequency analysis.

4.5 Summary

In this chapter, a method to solve the inverse map of the combined non-linear optical and fault models, was described. By solving this inverse map, two fault parameters $\beta_{\mathrm{f}}[n]$ and $\beta_{\mathrm{r}}[n]$ can be computed as well as an estimate of focus and radial distances $e_{\mathrm{f}}[n]$ and $e_{\mathrm{r}}[n]$. These four parameters parameterise the four dimensional detector space. $\beta_{\mathrm{f}}[n]$ and $\beta_{\mathrm{r}}[n]$ are transformed into two related residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$.
In Section 4.1 the overall solver structure is presented, where a Newton-Raphson based

iterative method is used to solve the inverse map, see Section 4.2, and a Kalman estimator is used to estimate the focus and radial distances, based on the solution to the inverse map, see Section 4.3. The potentials of the new residuals are illustrated by the fault detection example in Section 4.4.

Chapter 5

Time-Frequency Analysis

In the previous chapter a pair of decoupled residuals were derived, since the surface faults have clearly limited duration, it might be useful to use time-frequency analysis to extract some useful features from the residuals and the distance signals. An example of this is [Schneiders, 2001], where wavelets are used to detect surface faults on CDs. In order to do this time-frequency based time localisation and feature extraction, some theory is required. In this chapter the required theory is presented, before it is used for time localisation in Chapter 6 and feature extraction in Chapter 7. In Section 5.1 the back-ground for the time-frequency theory is given as well as a motivation for the usage of the four following bases. The wavelet basis is presented in Section 5.2, the wavelet packet basis is presented in Section 5.3, discrete cosine basis is presented in Section 5.4 and Karhunen-Loève basis is presented in Section 5.5.

5.1 Motivation of different time-frequency bases

The standard bases used for analysis of signals in the field of control engineering are: the Dirac basis (time basis) and the Fourier basis (frequency basis). (Since these analyses in practice are performed on finite signals, one needs to assume that this signal is periodic, and the analysed signal sequence, is multipla of periods of the signal.) These bases are not always well suited to represent the analysed signals. The Dirac basis has good time resolution, but no frequency resolution, on the other hand the Fourier basis has no time resolution, but a good frequency resolution, see [Hess-Nielsen and Wickerhauser, 1996]. These two bases are plotted in a time-frequency plane in the discrete case in Fig. 5.1 for the Dirac basis and in Fig. 5.2 for the Fourier basis.
If one uses these bases to analyse a fault can result in wrong conclusions. A surface fault like a scratch, has a small time duration compared to the time duration of one revolution. This means that even though a fault local in time introduces large levels of energy at certain frequencies, this can hardly be seen in a Fourier analysis of one

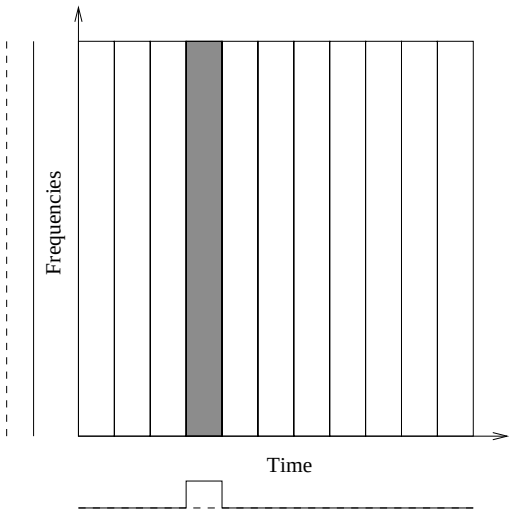


Figure 5.1: The time-frequency plane of the Dirac basis in the discrete case. Each of the blocks illustrates the projection of the basis vectors on the time axis, and their Fourier transforms projection on the frequency axis, illustrated by an example of the gray-scaled block.

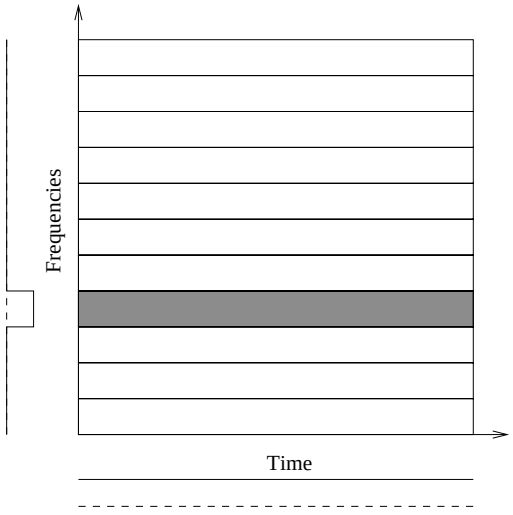


Figure 5.2: The time-frequency plane of the Fourier basis in the discrete case. Each of the blocks illustrates the projection of the basis vectors on the time axis, and their Fourier transforms projection on the frequency axis, illustrated by an example of the gray-scaled block.

or more revolutions. On the other hand if one performs the Fourier analysis of a time block which only contains the scratch and assumes it to be periodic with the length of the scratch as one period, these increased energy levels at certain frequencies can be seen. However, this kind of analysis removes all information about the frequencies of the remaining part of the revolution. If a time analysis was performed on the same signal, it would be impossible to separate low frequently variations in the signals (eccentricity and skewness) from a surface fault. If a block in time consisting of several samples is analysed, the frequency information in this block is needed in order to determine if a fault is present or not.

This makes it difficult to separate normal signals from faulty ones, and thereby detect the fault. It would be much more interesting to use a joint time and frequency analysis which both have time and frequency resolutions. One way to achieve this is by using a method which can give frequency information of time blocks not much larger than the scratch.

The best known of these joint time-frequency analysis methods is the Short Time Fourier Transform (STFT), where a Fourier transform is performed of a windowing function convoluted with the original signal, by assuming this block to be periodic, see [Mallat, 1999] and [Vetterli and Kovačević, 1995]. The window is following shifted in time such that the entire signal is covered by these windows. A time-frequency plot of this STFT basis can be seen in Fig. 5.3. A problem with the STFT is it introduces a boundary problem by the windowing of the signal, see [Mallat, 1999]. Another important thing one should notice about this figure, is the fact that even though this basis gives resolution in both time and frequency the time-frequency resolution blocks, called atoms, have the same shape in the entire time-frequency plane, see [Mallat, 1999] and [Hess-Nielsen and Wickerhauser, 1996]. In the analysis of a given signal it might be better suited to have different shapes of these atoms depending on their placement in the time-frequency plane. One should notice that, the sizes of the atoms always have a minimum, due to Heisenberg's uncertainty principle, see [Mallat, 1999] and [Wickerhauser, 1994a]. The wavelet transform is such a time-frequency based transform, which has time-frequency atoms of different shapes depending on the placement in the time-frequency plane, it has high frequency resolution at low frequencies and as consequence poor time resolution at low frequencies. The frequency resolution decreases and the time resolution increases as the frequency is increased, see Fig. 5.4. The wavelet transform has been used in many different applications: Detection and classification of faults in power delivery systems, [Mokhtari et al., 2002], [Parameswariah and Cox, 2002], [Xiangxun, 2002], and detection and classification of ECG signals: [Li et al., 1995] and [Dokur et al., 1999], fault detection: [Ye et al., 2004], [Ye et al., 2003], [Lopez et al., 1994] and [Ye et al., 2002], system identification for non-linear control: [Sureshbabu and Farrell, 1999], detection and feature extraction of geographic and meteorologic data: [Gamba and Marazzi, 1997] and [Ulfarsson et al., 2002], signal and image compression: [Strang and Nguyen, 1997], [Vetterli and Kovačević, 1995], and

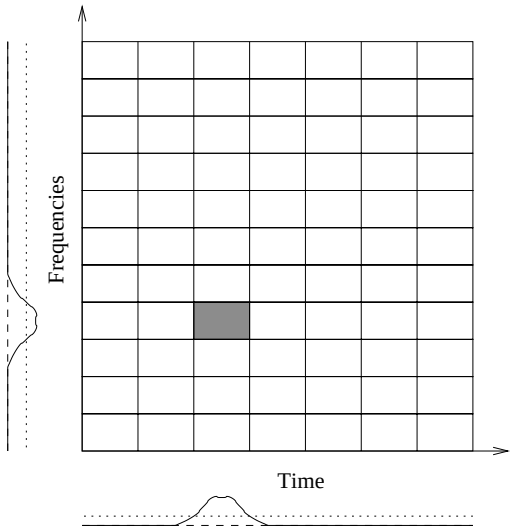


Figure 5.3: The time-frequency plane of the Short Time Fourier basis in the discrete case. Each of the blocks illustrates the location of most of the energy of the basis vectors on the time axis, and their Fourier transforms projection on the frequency axis, illustrated by an example of the gray-scaled block.

[Mallat, 1999], and rail fault detection: [Toliyat et al., 2003].

An extension of the wavelet transform, the wavelet packets transform, makes it possible to adapt the needed resolution in the time-frequency plane, with one important constraint: the area of the time-frequency atoms are constant. An example of such a wavelet packet transform time-frequency plane can be seen in Fig. 5.5. Another time-frequency transform which is of interest in this work is the block cosine transform, see [Mallat, 1999] and [Vetterli and Kovačević, 1995]. In the discrete version it is called Discrete Cosine Transform, DCT. This transform is much similar to the STFT, but it has less boundary effects than the STFT, see [Mallat, 1999]. The wavelet transform and the wavelet packet transform are designed to divide the frequency axis into intervals of varying sizes. These bases are well adapted to decompose signals that have different appearance in different frequency intervals. On the other hand the DCT is a block basis which separates the time axis. As a consequence the DCT is well suited to analyse signals which are well separated in time, see [Mallat, 1999].

The wavelet and wavelet packet transforms are interesting to use for analysing the signals containing the surface fault, since they might be useful to separate the surface fault signal component from the normal signal component, (due to noises and pick-up position). The DCT might be well suited to detect the location of the surface fault in time, since a surface fault presumably changes the frequency content of the analysed signal

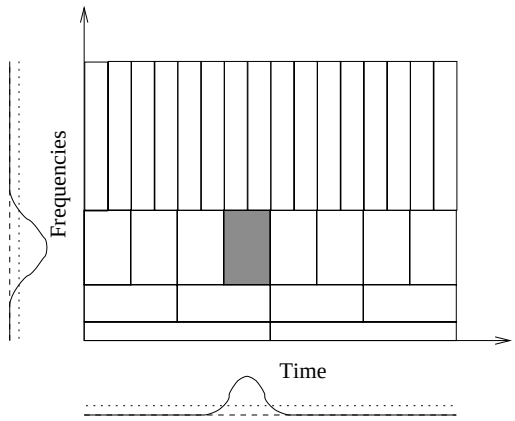


Figure 5.4: The time-frequency plane of a Wavelet basis in the discrete case. Each of the blocks illustrates the location of most of the energy of the basis vectors on the time axis, and their Fourier transforms projection on the frequency axis, illustrated by an example of the gray-scaled block.

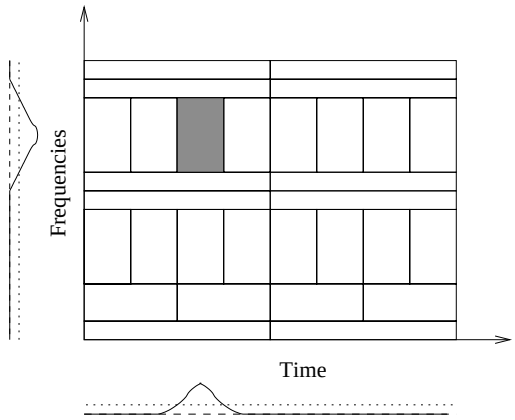


Figure 5.5: Principle time-frequency plane of an example of a Wavelet packet basis in the discrete case. Each of the blocks illustrates the location of majority of the energy of the basis vectors on the time axis, and their Fourier transforms projection on the frequency axis, illustrated by an example of the gray-scaled block.

almost instantaneously.
Another interesting basis for Time-Frequency based approximation of the surface faults is the Karhunen-Loève basis. Given a set of signals, the Karhunen-Loève transform of this set is derived based on the set and minimises the average linear approximation error, see [Mallat, 1999]. The Karhunen-Loève basis is related to the wavelet packet basis in the way that if one finds the wavelet packet basis in a special way, it can be proven that it is a fast computed approximation of the Karhunen-Loève basis, see [Wickerhauser, 1994a].
In the following the four bases: wavelet, wavelet packet, discrete cosine and Karhunen-Loève, are described in more details, with the purpose of getting the reader to understand the use of them in feature extraction of surface faults, which again is followed by a description of the basis.

5.2 Wavelet basis

The wavelet basis is an orthogonal basis of $\mathbf{L}^2(\mathcal{R})$, which is formed by scalings and translations of the mother wavelet, $\Psi(t) \in \mathbf{L}^2(\mathcal{R})$, with zero average:

$$\int_{-\infty}^{+\infty} \Psi(t)dt = 0. \tag{5.1}$$

The mother wavelet is normalised such that: $\left\|\Psi\right\| = 1$, and it is centred in the vicinity of $t = 0$. A family of orthogonal functions can be formed by scaling Ψ with s and translating it by u

$$\Psi_{u,s}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-u}{s}\right). \tag{5.2}$$

These new functions do remain normalised as the mother wavelet. The wavelet transform of $f \in \mathbf{L}^2(\mathcal{R})$ at time u and scale s is, see [Mallat, 1999]:

$$Wf(u,s) = \langle f, \Psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{(s)}}\Psi^*\left(\frac{t-u}{s}\right)dt. \tag{5.3}$$

$Wf(u,s)$ is only known for $s < s_0$ to recover f a complement to the information corresponding to $Wf(u,s)$ is needed for $s > s_0$. This is called the scaling function, Φ that is an aggregation of the mother wavelet at scale larger than 1, see [Mallat, 1999]. The low frequent approximation of f at the scale s is

$$Lf(u,s) = \langle f, \Phi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{(s)}}\Phi\left(\frac{t-u}{s}\right)dt. \tag{5.4}$$

This continuous wavelet transform is computationally heavy to use, due to the integral, whereas the Discrete Wavelet Transform, DWT, can be implemented by the use of filter banks. This transform is described and used by many authors, some examples on these are: [Mallat, 1999], [Jensen and la Cour-Harbo, 2001], [Strang and Nguyen, 1997], [Vetterli and Kovačević, 1995], [Wickerhauser, 1994a].

Discrete Wavelet Transform By defining the signal, which is intended to be transformed, as an approximation at scale 1, by a_1, the DWT decomposes successively each approximation at scale j into a coarser approximation at scale $j+1$, denoted a_{j+1}, and the remainder is denoted the details at scale $j+1$, denoted d_{j+1}.
The wavelet, $\Psi_{j,n}$, and the scaling function, $\Phi_{j,n}$, where j is the scaling and n is the time dilation. These functions form respectively bases of the approximation spaces and the detail spaces. The projection on these spaces is characterised by

$$a_j[n] = \langle f, \Phi_{j,n} \rangle \tag{5.5}$$
$$d_j[n] = \langle f, \Psi_{j,n} \rangle . \tag{5.6}$$

The projections in (5.5) and (5.6), can be represented by a convolution by two different filters. The approximation filter is denoted $\mathbf{h}$ and the detail filter $\mathbf{g}$, see [Mallat, 1999]. These filters are orthogonal quadrature filter pairs, if they fulfill the following conditions, [Wickerhauser, 1994a].

- Self-duality: $\mathbf{h}\mathbf{h}^* = \mathbf{g}\mathbf{g}^* = \mathbf{I}$;
- Independence: $\mathbf{g}\mathbf{h}^* = \mathbf{h}\mathbf{g}^* = 0$;
- Exact reconstruction: $\mathbf{h}^*\mathbf{h} + \mathbf{g}^*\mathbf{g} = I$;
- Normalisation: $\mathbf{h}\mathbf{1} = \sqrt{2}\mathbf{1}$, where $\mathbf{1} = \{\cdots, 1, 1, 1, \cdots\}$.

These conditions guarantee that it is possible to analyse the signal by using the wavelet transform. This analysis is called decomposition of the signal, since the transform decomposes the signal into time sequence blocks. The conditions also guarantee that from these blocks, the signal is decomposed into, it is possible to reconstruct the original signal. This process is called reconstruction of the signal.
Mallat proves the following Theorem in [Mallat, 1999] and [Mallat, 1989].

Theorem 5.1 (Mallat) *At the decomposition*

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] = a_j * \overline{h}[2p], \tag{5.7}$$
$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g[n-2p]a_j[n] = a_j * \overline{g}[2p]. \tag{5.8}$$

At the reconstruction

$$\begin{aligned} a_j[p] &= \sum_{n=-\infty}^{+\infty} h[n-2n]a_{j+1}[n] + \sum_{n=-\infty}^{+\infty} g[n-2pn]a_{j+1}[n] \\ &= \breve{a}_{j+1} * h[p] + \breve{d}_{j+1} * g[p] \end{aligned}, \tag{5.9}$$

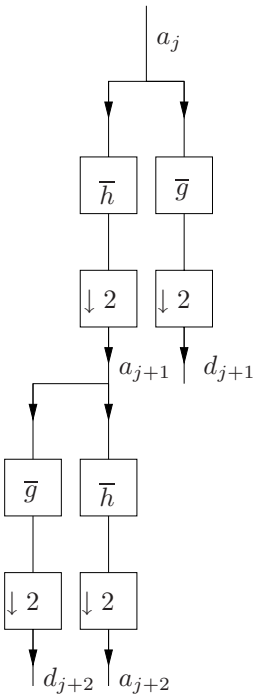


Figure 5.6: Illustration of the cascade filter implementation of the wavelet transform. $\overline{h}$ is the low-pass filter and $\overline{g}$ is the low-pass filter. a_j is the approximation at level j, and d_j is the details at level j.

where $\overline{x}[n] = x[-n]$ and $\breve{x}[n] = \begin{cases} x[p] & \text{if } n = 2p \\ 0 & \text{if } n = 2p+1 \end{cases}$. Notice the summation goes from $-\infty$ to ∞ even though the filters and signals in practice are finite. The theorem is still valid with finite signals, but some boundary effects can occur if the boundaries are not handled in an optimal way. The cascade filtering implementation of the DWT is illustrated in Fig. 5.6. This cascade filtering can continue as long as a_j can be decomposed, which means that if one would like to decompose a_j, the length of a_j is required to be larger than 2.

After this definition of the wavelet decomposition and reconstruction, some different wavelets functions are investigated. The choice of the wavelet is of high importance in the analysis of a signal. E.g. if the scope of an analysis is to detect a given signal component, it could be a scratch in the residuals. It is clearly a good idea to use a wavelet that has an appearance similar to the scratch. As a consequence of the importance of the wavelets on the achieved analysis results, there is defined a high number of different wavelets. In Fig. 5.7 the wavelet and scaling functions of the Daubechies 1-4 are

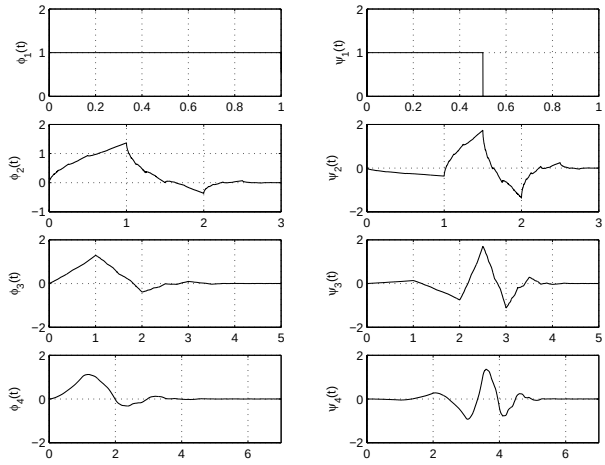


Figure 5.7: The mother wavelet and scaling function of the Daubechies 1-4 wavelets, where Daubechies 1 is also known as the Haar wavelet.

illustrated.
A problem with the orthogonal wavelets is that these wavelets are non-symmetric. Symmetric filters can be achieved if biorthogonal filters are used instead of the orthogonal ones. Biorthogonal bases consist of two sets where a vector in one set is orthogonomal on all the other vectors in the other set, but not on all vectors in the set they belong to, the biorthogonal filters are designed based on these biorthogonal wavelet bases. The biorthogonal filters consist of a pair of decomposition filters $(\mathbf{h}, \mathbf{g})$, and the reconstruction filters are the duals of these filters $(\tilde{\mathbf{h}}, \tilde{\mathbf{g}})$, see [Strang and Nguyen, 1997] and [Mallat, 1999]. The practical usage of the biorthogonal wavelet filters are the same as the orthogonal wavelet filters except for one difference. If the biorthogonal filters are used, another pair of filters is used to reconstruct the signal, than the filter pair used for the decomposition.

5.3 Wavelet packet basis

By viewing Fig. 5.6 one might ask, why d_{j+1} is not decomposed as a_{j+1}? Doing this leads to a decomposition tree illustrated in Fig. 5.8. It will result in a number of possible bases.
From a wavelet packet tree it is possible to choose among a number of different orthogonal bases. An example of a wavelet packet tree is a tree with three levels, (starting with level 0 to level 2). These possible bases are illustrated in Fig. 5.9.
The number of possible bases can be computed recursively, see [Wickerhauser, 1994a]. Let N_L be the number of bases in a wavelet packet tree with $1 + L$ levels. N_L can be

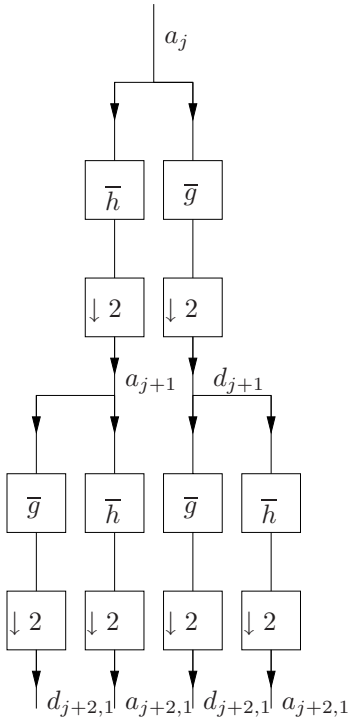


Figure 5.8: Illustration of the cascade filter implementation of the wavelet packet transform. $\overline{h}$ is the low-pass filter and $\overline{g}$ is the low-pass filter. a_j is the approximation at level j, and d_j is the details at level j.

computed recursively by the use of (5.10) and (5.11).

$$N_0 = 1, \tag{5.10}$$

and

$$N_{L+1} = 1 + N_L^2. \tag{5.11}$$

A simple estimate of N_{L+1} is $N_{L+1} > 2^{2^L}$ for all $L > 1$, [Wickerhauser, 1994a]. The number of possible bases for wavelet packet trees with 1-8 levels are shown in Table 5.1. These high numbers of possible bases for such small number of levels in the wavelet packet tree indicate the need for a fast algorithm to find the best basis in the wavelet packet tree, given a wavelet basis, information cost function and training signal. The best basis search is such a method, see [Wickerhauser, 1994a] and [Coifman and Wickerhauser, 1992].

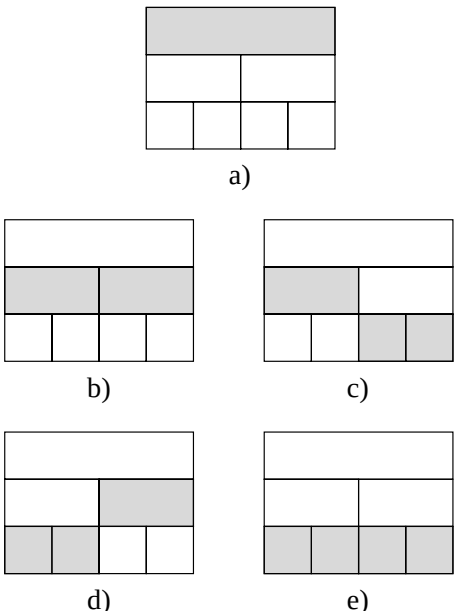


Figure 5.9: Illustration of the possible basis of a wavelet packet tree of 3 levels. The gray area illustrates the possible bases in the five wavelet packet trees. a) is the basis of the original signal, where b)-e) are the remaining four possible wavelet packet bases.

Number of levels in tree	Number of possible bases
1	1
2	2
3	5
4	26
5	677
6	458330
7	210066388901
8	44127887745906175987802

Table 5.1: The number of possible bases in the wavelet packet tree for 1-8 levels, see [Jensen and la Cour-Harbo, 2001]

Best basis search The best basis obviously depends on the application of which it should be found. In order to measure how suitable the basis is, an information cost function is introduced. Such a cost function measures the cost of a given representation, where the best basis has the smallest cost. The cost function, $\mathcal{K}(\cdot)$ must fulfill some specific properties.

1. $\mathcal{K}(\cdot) \in \mathcal{R}$
2. $\mathcal{K}\left(\begin{bmatrix}\mathbf{v} & \mathbf{w}\end{bmatrix}\right) = \mathcal{K}(\mathbf{v}) + \mathcal{K}(\mathbf{w})$, for finite length vectors $\mathbf{v}$ and $\mathbf{w}$.
3. $\mathcal{K}(\mathbf{0}) = 0$.

Some of the commonly used information cost functions are: Number above a given threshold, Concentration in l^p, Entropy, and Logarithm of energy.
Having the information cost function in mind, it is possible to describe the best basis search, [Wickerhauser, 1994a].

1. Compute the cost function of all elements in the wavelet packet tree, see Fig. 5.8.
2. Mark all elements on the bottom level J
3. Let $j = J$
4. Let $k = 0$
5. Compare the cost value v_1 of element k, (counting from the left), on level $j-1$ to the sum v_2 of the cost values of the elements $2k$ and $2k+1$ on level j.
   (a) If $v_1 \leq v_2$, all the marks below element k on level $j-1$ is deleted, and element k is marked.
   (b) If $v_1 > v_2$, the cost value v_1 of element k is replaced with v_2.
6. $k = k+1$. If there are more elements on level j(if $k < 2^{j-1}-1$), jump to step 5.
7. $j = j+1$. If $j > 1$, jump to step 4.
8. The marked basis has the lowest possible cost value. This value is fund at the top element.

The best basis search algorithm is illustrated by an example, which can be found in Fig. 5.10 where a signal of length 8 is decomposed, with the Haar basis, the decomposition tree can be seen in the top of the figure. Next the information cost function values are computed for all the elements, with the cost function: counting numbers in each element being absolute larger than 1. The algorithm is carried out for this example in the lower part of the figure. The use of this algorithm ends up with a best basis representation of the original signal being: $\begin{bmatrix}|10 & 6 & 22 & 6| & |-1| & |1| & |3 & 3|\end{bmatrix}$.

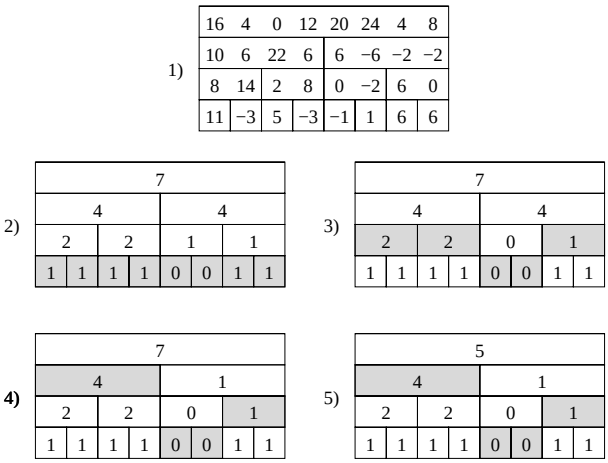


Figure 5.10: An illustration of the best basis search. The threshold above 1 is used as cost function and the Haar wavelet is the used wavelet, where 1) is the original wavelet packet tree, 2) is the computed informational costs, 3)-5) are the iterations in the best basis algorithm, ending up within a best basis representation of the signal being: [|10, 6, 22, 6|,|-1|,|1|,|3, 3|]

This best basis is computed based on one signal, cost function and wavelet. One can choose among the usable informational cost functions and wavelet bases. However, using only one signal to find the best basis is not always useful in practical applications. This can be illustrated by an example. The problem is to detect a given fault in a residual, N time sequences of measurements of this residual containing a fault is available. All these N time sequences of measurement are probably not all alike. This means that it could happen that the chosen signal for training the best basis, is a poor representation of other signals in the set of all the measured signals. It would be better to base the best basis search on more than one signal from the data set. Ideally the best basis search should be based on the entire data set.

Joint best basis The joint best basis is a method which takes the entire data set into account when finding the best basis. The joint best basis search finds the best basis given a set of signals, of the same length. It could be a number of signals with the same fault. The joint best basis algorithm computes the jointly best bases given: the set of signals, an information cost function and a wavelet basis, see [Wickerhauser, 1994a], [Coifman and Wickerhauser, 1992] and [Saito, 1998]. The algorithm is as follows:

1. Compute the full wavelet packet tree of all the signals in the signal set.

2. Compute the tree of means, by computing the mean of all signal trees at each position in the tree.
3. Compute the tree of squares, by computing the sum of squares of all signal trees at each position in the tree.
4. Subtract the tree of means from the tree of squares, to obtain the tree of variances.
5. Finding the best basis of the tree of variance by using the best basis algorithm, given an information cost function and wavelets.

The wavelet packet basis found by the use of the joint best basis is an approximation of the Karhunen-Loève basis, see [Wickerhauser, 1994a], which will be described in Section 5.5.

5.4 Discrete cosine basis

A block basis such as the discrete cosine basis is a product of two functions: a basis of an interval on the time axis, e.g. a cosine basis, and a windowing function which divides the time axis in consecutive intervals, as illustrated in Fig. 5.11. The upper part of the figure shows a separation in ideal windows, which are not practically usable. A practically used windowing function, the bell function, is illustrated in the lower part of the figure, see [Vetterli and Kovačević, 1995].
The cosine basis can be constructed in two ways, the cosine I and cosine IV basis. The differences between the two bases are due to the length of the period which the signal is extended into. Define $f[n]$ being the signal which is convolved with the windowing function. The cosine I transform, [Mallat, 1999], is

$$\hat{f}_I[k] = \lambda_k \sum_{n=0}^{N-1} f[n] \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right], \text{ with } \lambda_k = \begin{cases} 2^{-1/2} & \text{if } k=0 \\ 1 & \text{otherwise} \end{cases}, \quad (5.12)$$

and the cosine IV transform is

$$\hat{f}_{\mathrm{IV}}[k] = \sum_{n=0}^{N-1} f[n] \cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)\left(k+\frac{1}{2}+n\right)\right]. \quad (5.13)$$

The block cosine basis is illustrated in Fig. 5.12, where it can be seen that the windowing function shapes the amplitude of the cosine basis, see [Wickerhauser, 1994a].

5.5 Karhunen-Loève basis

Given a matrix, $\mathbf{X}$, of u column vectors in $\mathcal{R}^m$, where $u > m$, the Karhunen-Loève basis minimises the average linear approximation error of the vectors in the set,

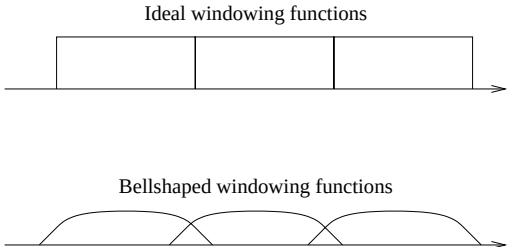


Figure 5.11: The upper part illustrates an ideal windowing function, where the lower part shows a bell shaped one, which is of the normally used ones.

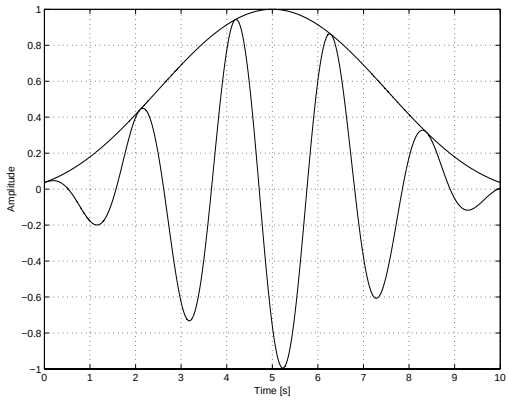


Figure 5.12: The local cosine and the window function.

[Mallat, 1999]. Another advantage of the Karhunen-Loève basis is that it approximates the general structures of all the signals in $\mathbf{X}$ with just a few basis vectors, meaning it can be used to compress the signals in $\mathbf{X}$, see [Mallat, 1999] and [Wickerhauser, 1994a] and [Saito, 1998]. The Karhunen-Loève basis has been used to a compression problem of pictures of faces, in [Wickerhauser, 1994b]. This data set of pictures of faces (128×128 pixels, eight bit gray scale with several thousand faces, see [Wickerhauser, 1994b]) constitutes a very large data set. The idea was to extract certain features from these pictures of faces, in order to identify one face among all the others. Dealing with feature extraction from a large size data set is problematic. Size of the data set was as a consequence reduced by the use of the Karhunen-Loève basis, [Wickerhauser, 1994b], without removing the general structures of the data, since these general structures are remaining. The important features can still be extracted from the data set, It is whereby still possible to identify a given face from the data set, see [Wickerhauser, 1994b]. In [Kraut et al., 2004] another usage of the Karhunen-Loève basis was reported, where it

was used for efficient estimations of tropospheric refractivity using radar clutter.
The Karhunen-Loève basis is computed based on $\mathbf{X}$, first of all it is assumed that the column vectors in $\mathbf{X}$ has zero mean, if not a preliminary step is introduced in order to fulfill that assumption. The Karhunen-Loève basis, $\mathcal{K}$, can be defined as

$$\mathcal{K} = \{v_1, \cdots, v_m\}, \tag{5.14}$$

is an orthornomal basis of eigenvectors of the matrix $\mathbf{X}\mathbf{X}^T$, ordered in such a way that v_n is associated with the eigenvector λ_n, and $\lambda_i \geq \lambda_j$ for $i > j$. A matrix of the basis vectors can following be defined as

$$\mathbf{K}_{\mathrm{L}} = \begin{bmatrix} v_m, v_{m-1}, \cdots, v_i \end{bmatrix} \tag{5.15}$$

So in other words the Karhunen-Loève basis is the eigen vectors of the autocorrelation of $\mathbf{X}$. The eigenvalues of the autocorrelation have the values of the variances of the related Karhunen-Loève basis vectors. The approximating properties of the Karhunen-Loève basis vectors are sorted in increasing order, that means that if the basis consists of p vectors the basis vector p is the most approximating basis vector. In addition the general structures in all the vectors in $\mathbf{X}$ are represented by only a few basis vectors. The remaining basis vectors represent the signal parts which are not general for $\mathbf{X}$, i.e. noises in the signal, etc.
It can be proven that

$$\mathbf{X} = K_{\mathrm{l}} \Sigma L_{\mathrm{l}}, \tag{5.16}$$

where

$$\Sigma = \begin{bmatrix} \tilde{\Sigma} & 0 \\ 0 & 0 \end{bmatrix}, \tag{5.17}$$

$$\tilde{\Sigma} = \text{diagonal}(\sigma_1, \cdots, \sigma_p), \tag{5.18}$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0, \tag{5.19}$$

and L_{l} is a unitary matrix. From this it can clearly be seen that this decomposition is the Singular Value Decomposition.

5.6 Summary

In this chapter the necessary back-ground theory in the field of time-frequency analysis has been described, starting with a general description of the field of time-frequency analysis, see Section 5.1. This is followed by the description of four different interesting bases. The wavelet basis is described in Section 5.2, and the wavelet packet basis is described in Section 5.3, where both the best basis search and the joint best basis search are described. The discrete cosine basis is described in Section 5.4, which describes both DCT-I and DCT-IV. The Karhunen-Loève basis is described in Section 5.5.

Chapter 6

Time Localisation of the Surface Faults

After the introduction to time-frequency based bases in Chapter 5, and the construction of decoupled residuals in Chapter 4, the increased fault detection powers of these residuals have been illustrated in Section 4.4. These bases can be useful in the time localisation of the surface faults. The time location of the surface faults is found based on analysis of the residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$. The next problem is to locate the surface faults (fault detection) in time based on the residuals. Thresholding is often used to detect and locate the fault in the residual. [Basseville and Nikiforov, 1993] present methods to compute the thresholds based on statistical information. Schneiders has used wavelet analysis to detect surface faults in optical disc players, see [Schneiders, 2001] as well as [Goossens and Odgaard, 2003], however, inpsecting these one can come to the conclusion that it would be better to use wavelet packet fitlers instead. In this chapter some other methods are used: detection of frequency change, extended thresholding, and wavelet packet filters.

However, before these fault time localisation methods are described, some online computational remarks are given regarding the application, see Section 6.1.

The frequency content of the residuals change during surface faults. This makes it obvious to use a method to detect and locate the fault, which is based on time-frequency analysis, instead of the statistics of the residuals with and without faults. A time-frequency based method in this case might be able to detect the beginning of a fault from the background noise. This detection of the change in frequency content is similar to the vocal segmentation problem described in [Wesfreid and Wickerhauser, 1999], where Fang's algorithm for segmentation of the time axis was used, [Fang, 1994] and [Wesfreid and Wickerhauser, 1999]. This method is described in Section 6.2

On the other hand if a thresholding method is used, see Section 6.3, it needs to be adapted to some applicational problems. Some additional problems are encountered when sur-

face faults shall be detected. Signal noises are different from one disc to another, as a consequence the thresholds need to be adapted to the given noise level. The skewness of the disc influences the residuals by changing the received fraction of energy in the OPU. This eccentricity part is mainly due to the angular velocity of the CD-player, which results in a low frequent component in the residuals. It is also important to have a good localisation of the beginning and end of the faults. From experiences it is known that a poor fault localisation can result in severe problems for the controllers. Finally, it is of high importance to avoid false positive detections of faults, since the fault accommodation will somehow make the controllers non sensible to the frequency content of the faults. This means that the controllers cannot suppress disturbances in that frequency range during a detection of a fault. A consequence of that is that the time duration of the fault accommodation shall be reduced as much as possible.
The final method used to locate the fault in time, see Section 6.4, is wavelet packet filters designed partly based on the joint best basis search, where the best wavelet packet basis is found, based on a training set of measurements of real surface faults.

6.1 Online computation

It is problematic to use advanced signal processing online in a control system, since output of this signal processing is required at the time the fault is occurring. However, a surface fault such as a scratch has only small variations from an encounter to the next encounter, this is due to the small distance from track to track, (the distance is only 1.6 [μm]). This fact has been validated by experimental work, in [Odgaard et al., 2004d]. $\beta_{\mathrm{f}}[n]$, (the focus fault parameter, see (3.59)), of some sequently following encounters of a fault were plotted, and this plot clearly validates the point that a fault has a small variation from encounter to encounter. This can be seen from Fig. 6.1 and Fig. 4.2.
This means the time-location and features extracted from the surface fault at an encounter can be used when the same fault is occurring again at the next encounter. I.e. notice the two following important remarks: first of all the computations in the feature extraction can be spread out in longer time period and thereby make use of more advanced algorithms possible, since these algorithms are too computationally demanding to compute them at once. It is at least important to have the entire fault available when the time-location and features are extracted. It would be difficult, if not to say impossible to extract the frequency content of a fault based on just a few samples, or to classify the fault based on a couple of samples. It is a requirement that the time-location and features are instantly available if they are used to handle the faults in a closed loop control scheme. This can guaranteed if the time localisation from the last encounter of the fault is used together with a prediction of the length of one revolution.
Even though surface faults have small variations from encounter to encounter they do vary over a larger number of encounters. This means that surface faults are small in the beginning and being no problem to handle by the nominal controller, and develop over a large number of tracks to be a large fault, which is problematic to handle by the nominal

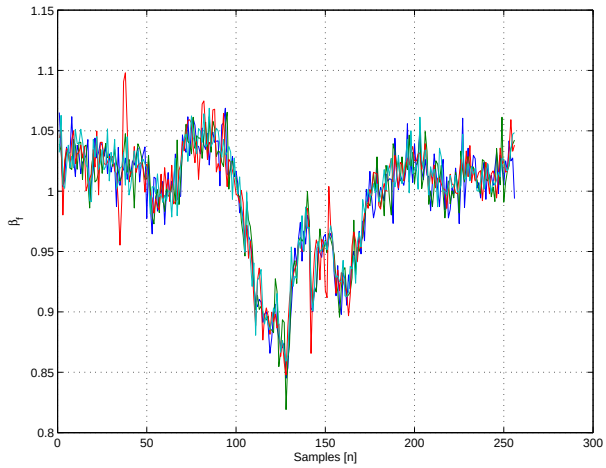


Figure 6.1: $\beta_f[n]$ of a small scratch from a number of encounters of the same fault. Notice the small variation from encounter to encounter.

controller. Some experimental works could indicate that many scratches have an eye shape, as illustrated in Fig. 6.2.
To validate this eye shape, a 3-dimensional plot of a scratch on a disc is plotted in Fig. 4.1.
In this figure $\beta_f[n]$ is plotted with a number of sequential following fault encounters, the depth and duration of the fault increase from encounter number 1 to the last encounter.

6.2 Time localisation based on Fang's segmentation

Fang's algorithm, [Fang, 1994], has not been published by its inventor, but it has been presented by other authors e.g. [Wesfreid and Wickerhauser, 1999]. This algorithm is well suited to locate changes in the frequency content of a signal. This algorithm consists of three parts:

- The instantaneous frequency change function at a given sample, n, $IFC(n)$.
- An averaged frequency change function at the sample, n, $AFC(n)$ which is an averaged version of $IFC(n)$.
- A locator of the change in frequencies. The change in frequencies can be located by finding local extrema in the $AFC(n)$ signal.

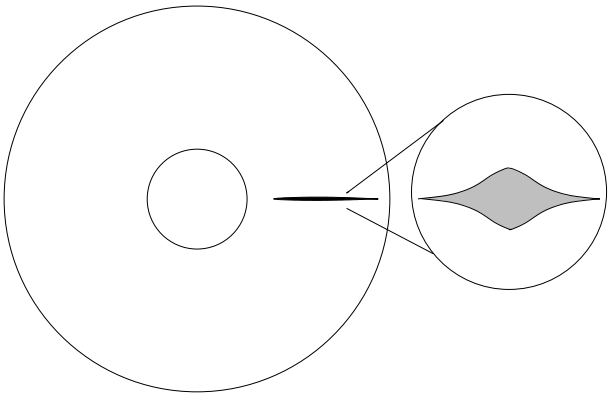


Figure 6.2: Illustration of the eye shaped scratch.

Instantaneous frequency change function The instantaneous frequency change function gives a measure of the change in spectrum of a given signal, $y[n]$. It is done by computing the change in flatness of the spectrum of the signal. Flatness can be defined in a way that a spectrum with uniform energy distribution has a high degree of flatness, and a spectrum with concentrated energy distribution has a low degree of flatness.
First define three intervals: $a_n : [n-l, n+l]$, $b_n : [n-l, n]$ and $c_n : [n, n+l]$, where $l \in \mathcal{Z}^+$. Now compute the flatness of the spectra of y in these three intervals: $\lambda_{a_n}(y)$ the flatness of spectrum of y in the interval a_n, $\lambda_{b_n}(y)$ the flatness of spectrum of y in the interval b_n and $\lambda_{c_n}(y)$ the flatness of spectrum of y in the interval c_n. The instantaneous frequency change function can now be defined ($IFC(n)$), see [Wesfreid and Wickerhauser, 1999]

$$IFC(n) = \lambda_{c_n}(y) - (\lambda_{a_n}(y) + \lambda_{b_n}(y)), \tag{6.1}$$

where $n \in \{\eta + l, \cdots, N - \eta - l\}$, η is the window overlap of the used spectrum transform, and N is the block length.
To explain how this method works, an example is illustrated in Fig. 6.3. The signal in this figure has step in the middle. Three intervals are defined on signal as shown figure. The spectrum of intervals a_n and b_n do only contain a DC-component, and the spectrum of the interval c_n has energy at all frequencies do to the step in the signal, which results in a flat spectrum. This means that $\lambda_{c_n}(y) > (\lambda_{a_n}(y) + \lambda_{b_n}(y))$. On the other hand if intervals shifted such that the step was placed in either a_n or b_n. The spectrum of the interval with the step would be more flat than the other spectrum, meaning that the IFC would a smaller value. I.e. the $IFC(n)$ will have its maximum if n is placed at step.
Until now a spectrum was assumed but the transformation used is not mentioned. The smooth DC4-IV transform is chosen since it is a smooth orthogonal basis with a good resolution of the frequency axis. In addition it has a fast implementation, see [Mallat, 1999]. The flatness of the spectrum can be measured by

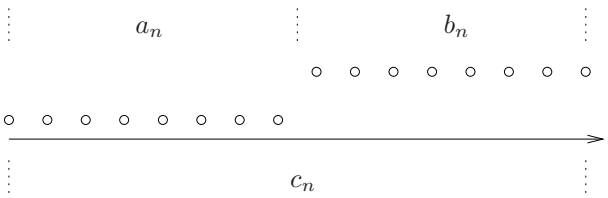


Figure 6.3: An example on Fang's algorithm for segmentation of the time axis. The three intervals a_n, b_n and c_n are defined related to the signal. The spectrum of intervals a_n and b_n do only contain a DC-component, and the spectrum of the interval c_n has energy at all frequencies do to the step in the signal, which results in a flat spectrum. This means that $\lambda_{c_n}(y) > (\lambda_{a_n}(y) + \lambda_{b_n}(y))$.

the Shannon entropy, (see 5.3 and [Mallat, 1999]), of the squared spectrum, see [Wesfreid and Wickerhauser, 1999]

$$\lambda(x_0, x_1, \cdots, x_n) = -\sum_{k=0}^{j-1} |x_k|^2 \log(|x_k|^2), \tag{6.2}$$

where $x_i,\ i \in [0 \cdots j]$ is the coefficients of the spectra.

Averaged frequency change function The IFC oscillates even when the signals are periodic, see [Wesfreid and Wickerhauser, 1999]. The IFC is as a consequence low-pass filtered to obtain the averaged frequency change function $AFC(n)$. This filter is made as follows. If h is a biorthogonal low-pass filter, see Section 5.2, and $\tilde{h}$ its dual, then

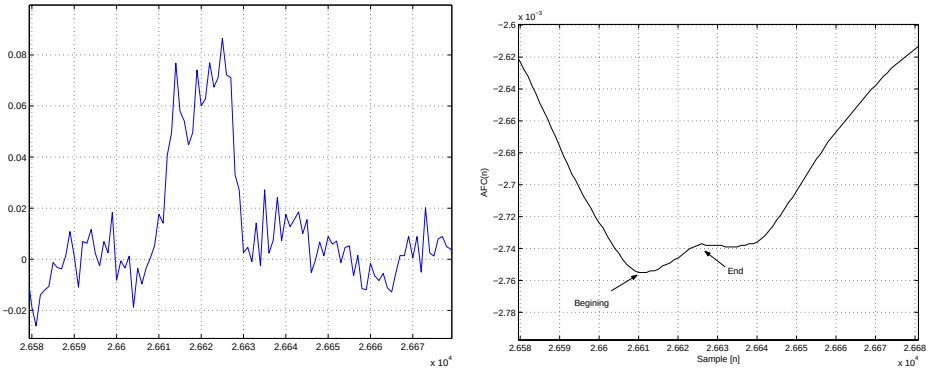
$$AFC(n) = \tilde{h}^d h^d (IFC(n)), \tag{6.3}$$

where $h^d = hhh \cdots h$, d factors. In this work the symmetric biorthogonal filters, used in the FBI/Los Alamos/Yale "WSQ" algorithm, is used. These filters are chosen due to their good symmetric properties, see [Ouyang and Wickerhauser, 1993]. The filtering with both h and its dual $\tilde{h}$ removes any phase shifts of the output of the algorithm.

Locator of frequency changes The localisation of the frequency changes, which in [Wesfreid and Wickerhauser, 1999] is called segmentation of the time axis. In [Wesfreid and Wickerhauser, 1999] the frequency changes are located by finding the local extrema in $AFC(n)$. These local extrema are found by detecting zero crossings of the adjacent differences of $AFC(n)$.

Fault detection based on AFC In the following the potential of the AFC method is compared with the normally used threshold method. The parameters of the AFC are

found by iterations, in a way that filters are focused on the changes in the frequency content of the fault, and not on the general frequency content of the signal. The used parameters are: $\eta = 16$, $l = 32$ and $d = 4$. The two first parameters primarily determine the location of the frequency region and the last parameter determines the width of this frequency change detection.
The comparison of the algorithms is done by using computed residuals from three data sequences with representative faults, see Figs. 6.4-6.6.

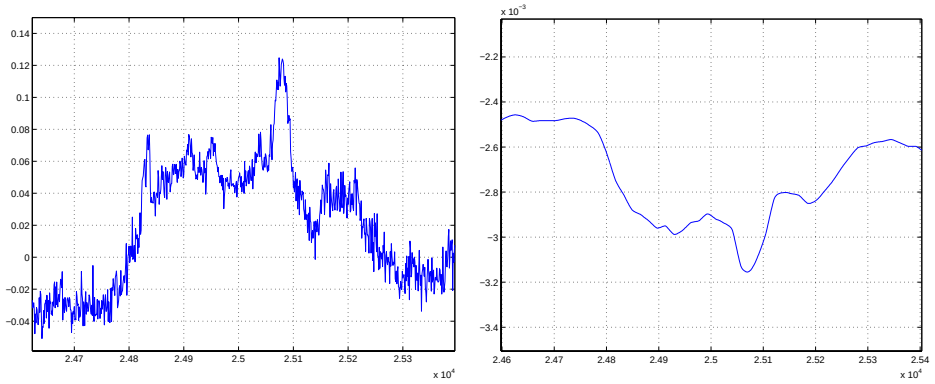


(a) The residual with first fault.

(b) The $AFC(n)$ of the residual with the first fault.

Figure 6.4: An illustration of the first fault in the residual and the $AFC(n)$ of the residual. The abscissis is in samples. Note the markings of the beginning and end of the scratch. These locations of the beginning and the end of the fault based on $AFC(n)$ are very suitable.

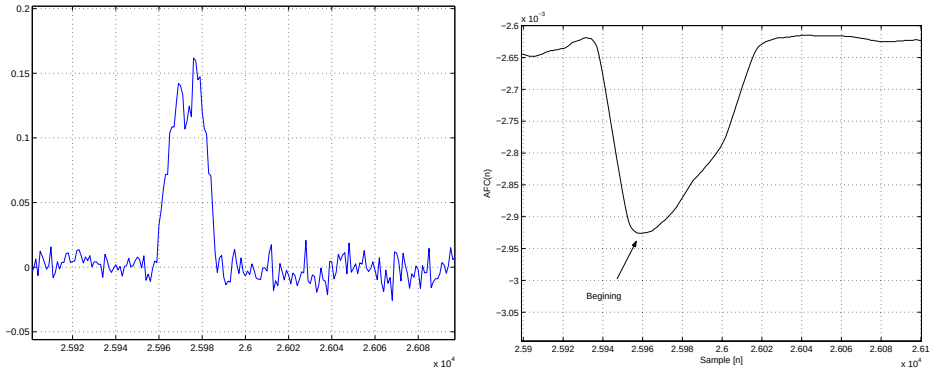
By inspection it can be seen that finding the local minima, (due to a sign difference compared with the original algorithm), of the AFC in Fig. 6.4(b), gives very good indication of the beginning and end of the fault illustrated in Fig. 6.4(a). It is also seen that the AFC in Fig. 6.5(b) cannot be used to locate the start and end of the fault, since its minimum does not locate the beginning and end of the fault. This means that the change in frequencies in the signal is not measured by the AFC function since the frequencies changing are not in the interval of frequencies for which the algorithm is designed to measure changes. Instead, the AFC in this example looks like a low-pass version of the residual itself, and could be time located by the use of a simple threshold method. On the other hand if the design is based on the fault in Fig. 6.5(a), it results in a design which cannot be used to detect the beginning and end of the fault in Figs. 6.4(a) and 6.6(a). A design based on the fault in Fig. 6.6(a) was also tried, with the result that it was well suited to locate the fault in Fig. 6.6(a) but not the fault in Figs. 6.5(a) and



(a) The residual with second fault.

(b) The $AFC(n)$ of the residual with the second fault.

Figure 6.5: An illustration of the second fault in the residual and the $AFC(n)$ of the residual. The abscissis is in samples. Note the lack of markings of the beginning and end of the scratch. This means that this $AFC(n)$ cannot be used to locate the faults in time.



(a) The residual with third fault.

(b) The $AFC(n)$ of the residual with the third fault.

Figure 6.6: An illustration of the third fault in the residual and the $AFC(n)$ of the residual. The abscissis is in samples. Note the markings of the beginning of the scratch. The $AFC(n)$ can in this example only be used to locate the beginning and not the end of the fault and is as a consequence not usable in the given situation.

6.4(a). The last example, Fig. 6.6(a), looks like the first one, Fig. 6.4(a), but the AFC of this signal has only one minimum, see Fig. 6.6(b), even though this minimum locates the start of the fault, this signal cannot be used to detect the end of it, since there is only one extrema in this signal.
This means that using the AFC of the residuals in some cases gives a much better time localisation of the fault than by using thresholds since the thresholds have to be larger than the signal noise. As a consequence the beginning of the fault cannot be localised well. The AFC of the first fault seems to do exactly this. Unfortunately it does not have the same good results in the other examples, for the second example, see Fig. 6.5(a). The AFC looks like a low-pass filtered version of the original example. The fault could be located by using a threshold. However, the DCT used in the AFC is a computationally demanding algorithm. The design of the AFC algorithm results in an over trained filter, which is too narrow banded to localise other surface faults. This problem is also mentioned in [Ye et al., 2004] and [Ye et al., 2003], where too narrow banded parity filters used for fault detection results in a high level of non-detections.
One way to use this method for time localisation of faults is to use a previous classification of the faults, and adapt a number of AFC filters to the different classes of faults. This is usable since the faults slowly but steadily are evolved for each revolution of the disc from being no problem to handle to be a problem, meaning it is possible to classify the fault before using the AFC filters to locate the fault in time.

Summary of time localisation by the use of Fang's algorithm This means that even though there is a clear potential of the AFC based detection for some faults see Fig. 6.4(a), it do not shows the same good potentials in situations like in Fig. 6.5(a). It cannot be concluded that using one AFC filter is recommendable in general for time localisation of surface faults on optical discs, since the required filters are too narrow banded for locating all the surface faults. A strategy could be to implement a number of AFC filters, where each is adapted to a given class of faults. This strategy would probably have an improved time localisation of the surface faults, but would require a high amount of computations.

6.3 Cleaning of residuals and extended threshold

Cleaning of residuals Skew discs introduces a problem if thresholds are used for the time localisation of the surface fault. The skewness of the disc results in oscillating references to the focus and radial distances, which are handled by the controllers. The used residuals are designed in a way such that they should be decoupled from these distances. However, in addition to these variations the skewness also results in oscillations in the received amount of energy at the detectors. This can be seen as oscillations in the residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$. The skewness is illustrated in Fig. 6.8, from this figure it is seen that a skew disc does not reflect all the light back to the OPU. A couple of residuals with a clear skewness problem is illustrated by an example from a disc with a scratch

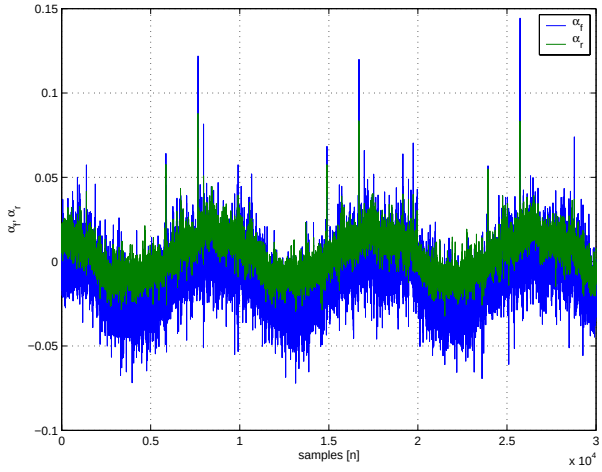


Figure 6.7: The two residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$ computed for a disc with a scratch and a skewness problem.

and a skewness problem, Fig. 6.7. In the following an "ad-hoc" method for handling this problem is suggested. If threshold based detection is performed on these residuals the required threshold for detection of the faults is dependent on the faults location on the disc, and if the fault is placed at the top of the oscillation this threshold needs to be high to ensure no false detections. This means a late detection.
The skewness component in the signal is low frequent, almost quasi constant. It can be removed at sample n by subtracting the mean of the block of samples $[n - 2^k, n - 1]$, where k is chosen in a way that the block is not too short or long, this gives a low frequent bandstop filter. Using this method the skewness components are removed from the example, see Fig. 6.9. An alternative to this approach is to use a notch-filter which is needed to be tuned to frequency of these low frequently components.

Extended threshold This mean of the block of samples will change during a fault. In order to avoid that, these mean blocks need to be much longer than the fault itself. This is not an optimal solution, since a long mean block requires either a large memory or a large number of computations. By inspection it is seen that the skewness component does not change during faults. This means that one "ad-hoc" solution to the mean problem is to fix the mean from the sample at which the fault is detected, to the sample one block length after the end of the fault. The meaning blocks are implemented in (6.6) and (6.8). Noises in $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$ make it difficult to use an absolute threshold, since the level of these noises varies. Instead a threshold relative to the variance of the non-fault residual parts can be used.
To make sure that it is a fault which is detected, the used threshold for the detection of the

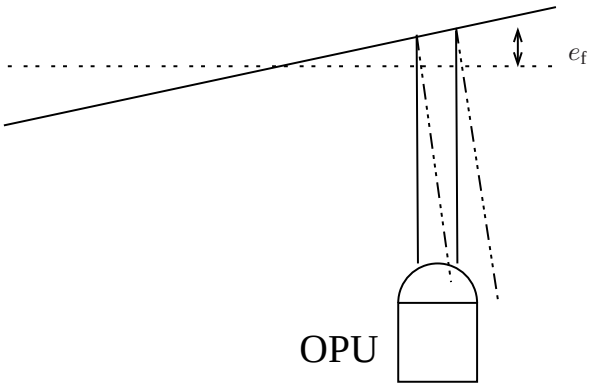


Figure 6.8: Illustration of the skewness of the disc. Notice that not all the reflected light reaches the OPU.

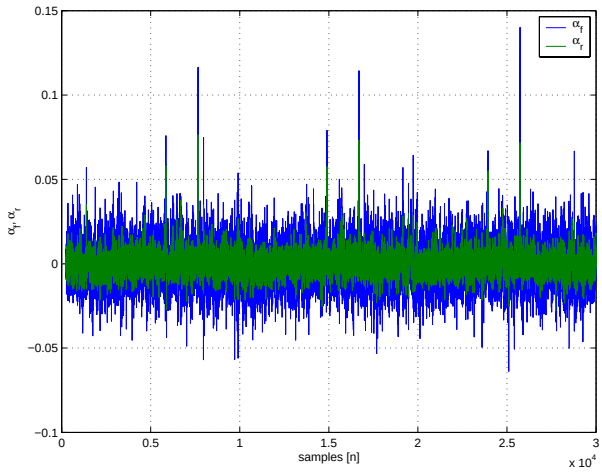


Figure 6.9: The skewness component is removed from the two residuals $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$ computed for a disc with a scratch and a skewness problem.

beginning of the fault needs to be larger than the noises in the residuals. However, having the knowledge of a fault present it is possible to have a lower threshold for detection of the end at the fault. In other words a multiple threshold is used. This means that using this multiple threshold, somehow solves two problems. The large threshold for beginning detection, γ_{beg}, limits the numbers of the false positive detections of faults. The small threshold for end detection, γ_{end}, improves the detection of the end of the fault, implemented in (6.5) and (6.7).
The mapping from the two residuals to one detection signal, is proposed in [Odgaard et al., 2003a], simply by taking the ∞ norm of the two detections. The detection decision schemes can now be described

$$d[n] = \left\| d_{\text{f}}[n] \quad d_{\text{r}}[n] \right\|_{\infty}, \tag{6.4}$$

where

$$d_{\text{f}}[n] = \begin{cases} 1 & \text{if } \frac{\tilde{\alpha}_{\text{f}}[n]}{\text{var}(\alpha_{\text{f}}))} > \gamma_{\text{beg}} \wedge d_{\text{f}}[n-1] = 0 \\ 1 & \text{if } \frac{\tilde{\alpha}_{\text{f}}[n]}{\text{var}(\alpha_{\text{f}}))} > \gamma_{\text{end}} \wedge d_{\text{f}}[n-1] = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{6.5}$$

$$\tilde{\alpha}_{\text{f}}[n] = \alpha_{\text{f}}[n] - \text{mean}(\alpha_{\text{f}}[n-2^k, n-1]), \tag{6.6}$$

and

$$d_{\text{r}}[n] = \begin{cases} 1 & \text{if } \frac{\tilde{\alpha}_{\text{r}}[n]}{\text{var}(\alpha_{\text{r}}))} > \gamma_{\text{beg}} \wedge d_{\text{r}}[n-1] = 0 \\ 1 & \text{if } \frac{\tilde{\alpha}_{\text{r}}[n]}{\text{var}(\alpha_{\text{r}}))} > \gamma_{\text{end}} \wedge d_{\text{r}}[n-1] = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{6.7}$$

$$\tilde{\alpha}_{\text{r}}[n] = \alpha_{\text{r}}[n] - \text{mean}(\alpha_{\text{r}}[n-2^k, n-1]). \tag{6.8}$$

Since the time localisation is not performed on-line, it is possible to improve the localisation of the beginning of the fault. The relative high threshold used for detection of the beginning is used, to avoid false detections, meaning localisation of none faults. But if a fault beginning is located, the algorithm could be extended by stepping back through the residuals as long as the residuals have a higher value than a threshold smaller than γ_{beg}, this smaller threshold is denoted $\gamma_{\text{beg,low}}$. This method forms the second version of extended threshold method.

6.3.1 Experimental results of extended threshold

In the following the two extended threshold methods are compared with a traditional threshold on some test data, which are different from data used in the AFC detection since these signals need to have a skewness component to show the potential of the extended threshold. The used experimental setup is described in Chapter 3.1.

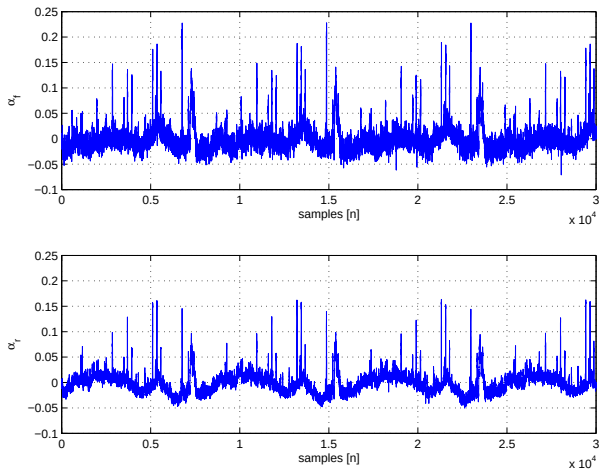


Figure 6.10: $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$ computed of sampled signals with surface faults and eccentricity problems.

Results on the extended threshold method Residuals of the test data are presented in Fig. 6.10.
Both the simple and the extended threshold methods were used for four different faults in the residuals illustrated in Fig. 6.10. The thresholds are found such that false positive detections are avoided, meaning that standard and first extended version methods should give almost the same beginning detection but both extended methods should give a better detection of the end of the fault, and the second version of the extended method should also improve the localisation of the beginning of the fault. The results are shown in Table 6.1. From the results in Table 6.1 it can be seen that both extended threshold methods perform better in the localisation of the end of the faults, which is expected due to the multiple thresholds. The second version of the extended threshold method does also improve the localisation of the beginning of the fault, whereas the first version of the extended method performs as well as the normal threshold. The potential of the extended method regarding removal of the eccentricity is only revealed in the last example. In this example both extended methods localise the beginning and end of the fault based on $\alpha_{\mathrm{f}}[n]$ close to the beginning found by inspection. The normal threshold method has a larger general locating error.

6.3.2 Summary of time localisation by extended threshold

The threshold method is extended to improve its handling of disc eccentricity and to use multiple thresholds, meaning that one is used for detection at the beginning and a smaller one is used for detection at the end. This gives a better time localisation at the

α_f, α_r	Inspection	Normal	Extended v.1	Extended v.2
α_f	2830-2870	2843-2865	2844-2867	2833-2867
α_r	2830-2870	2837-2863	2837-2870	2834-2870
α_f	6752-6780	6754-6776	6754-6778	6754-6778
α_r	6752-6780	6757-6773	6757-6775	6755-6775
α_f	13202-13227	13204-13223	13204-13227	13203-13227
α_r	13202-13227	13205-13222	13205-13224	13204-13224
α_f	13424-13470	13445-13465	13427-13469	13426-13469
α_r	13424-13470	13446-13464	13446-13465	13445-13465

Table 6.1: The four examples of the time localisation based on the two threshold methods, and the localisation by visual inspection.

end of the surface fault. Since the time localisation of the faults is performed off-line, the second version of the extended threshold is implemented by using a small beginning threshold, if a fault has been located.

6.4 Wavelet packet based time-localisation of the surface faults

It might be a good idea to use the wavelet and wavelet packet bases to locate the surface faults in time. This has been done, e.g. see [Schneiders, 2001], who uses wavelets to detect the surface faults, and in a different setup, the DVD-player, a wavelet packet based method is used for fault detection, which in principle is time localisation of the fault. This fault detection is a part of method to handle surface faults, see [Goossens and Odgaard, 2003]. Both methods are based on the normally used residuals, (the sum signals) and not the decoupled residuals derived in this thesis. The wavelet and wavelet packet filters designed in [Schneiders, 2001] and [Goossens and Odgaard, 2003] are based on only one training signal. This can result in a problem. If another fault is applied to the filter it does clearly not detect this fault as well as the surface fault used for the design of the filter.
This over-training to the faults in the training set results in wavelet filters which have a too narrow-pass band, meaning that the main energy of the other faults are outside the pass band of the designed wavelet filters. This is more or less the same conclusion obtained in [Ye et al., 2004] and [Ye et al., 2003], where it is concluded that a too narrow banded filter used for fault detection results in a clear increase in the non detected faults. This problem of designing a wavelet packet filter to detect surface faults which is not entirely alike, can be handled better by using the joint best basis search, instead of the best basis search, which only finds the best basis given one signal and not a number of

signals.

Training data From the focus and radial residuals, $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$, faults are extracted based on the algorithm described in [Odgaard and Wickerhauser, 2004]. Each detected fault is extracted into a column vector with the length of 256=2^8 samples. This length is chosen since all faults in the dataset are shorter than 256 samples. The faults are extracted with symmetric geometric centre intended to be in the middle of the fault vector. Each vector can contain several faults (especially fingerprints). In addition the centralisation is not totally successfully due to implementation. It was chosen that a given fault is only contained in one fault vector. I.e. the centre is not always in the middle of the vectors.
This extraction gives two matrices with faults. The faults in $\alpha_{\mathrm{f}}[n]$ are collected in $\mathbf{F}_{\mathrm{f}}$ and the faults in $\alpha_{\mathrm{r}}[n]$ are collected in $\mathbf{F}_{\mathrm{r}}$, where each column in the matrices is a fault vector.

Joint best basis A joint best basis is following found for $\mathbf{F}_{\mathrm{f}}$ and $\mathbf{F}_{\mathrm{r}}$, where some different wavelets also are tried. The problem in using the joint best basis algorithm directly is the choice of information cost function and the best basis search. Then used to fault detection, the scope is to find a band-pass filter which separates the surface fault from the background noises in the residuals. Instead, a heuristic based method is used. The method takes its starting point in the tree of variances, used in the joint best basis algorithm, see [Odgaard et al., 2004e]. The used method consists of the following steps:

1. Compute the tree of variances.
2. Search down the levels in the tree to find a level where the approximations and details both contain energy.
3. Use these details for the fault detection.

This method is following used on $\alpha_{\mathrm{f}}[n]$ and $\alpha_{\mathrm{r}}[n]$. Starting with $\alpha_{\mathrm{f}}[n]$, the Haar wavelet is used, since it is well suited for detection changes in the signal, and is short. By using the heuristic method, the interesting part of the tree of variance can be seen in Fig. 6.11. The figure starts with the original signal. The remaining figure parts are located by denoting a low-pass filtering with h and high-pass filtering with g, meaning that two low-pass filterings followed by one high-pass, are denoted hhg. The second plot is h, the third is hh, the fourth is hhh, the fifth is $hhhh$ and the last one is $hhhg$. Notice the large change from hhh to $hhhh$, which results in a significant signal in $hhhg$. This signal is useful for fault detection, since it has relatively large signal parts, and does not contain the near zero frequencies, where disturbances are dominating.
This means that a useful filter for fault detection in $\alpha_{\mathrm{f}}[n]$ is found. It is three Haar low pass filters followed by one Haar high-pass filter. The frequency amplitude plot of this filter is illustrated in Fig. 6.12.

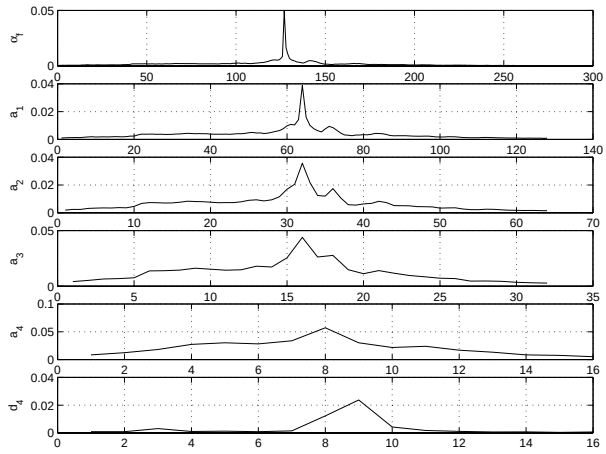


Figure 6.11: The part of the tree of variance of $\alpha_{\mathrm{f}}[n]$, which is interesting in the used method. The figure starts with the original signal. The remaining figure parts are located by denoting a low-pass filtering with h and high-pass filtering with h, meaning that two low-pass filterings followed by one high-pass, are denoted hhg. The second plot is h, the third is hh, the fourth is hhh, the fifth is $hhhh$ and the last one is $hhhg$.

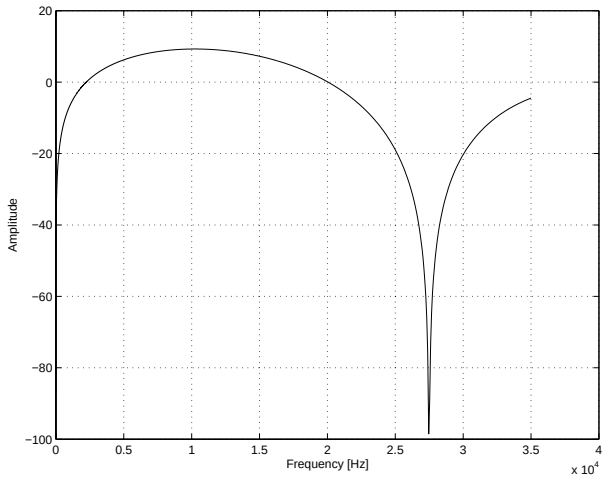


Figure 6.12: The frequency amplitude plot of filter consisting of three times Haar low-pass filtering followed by one time Haar high-pass filtering.

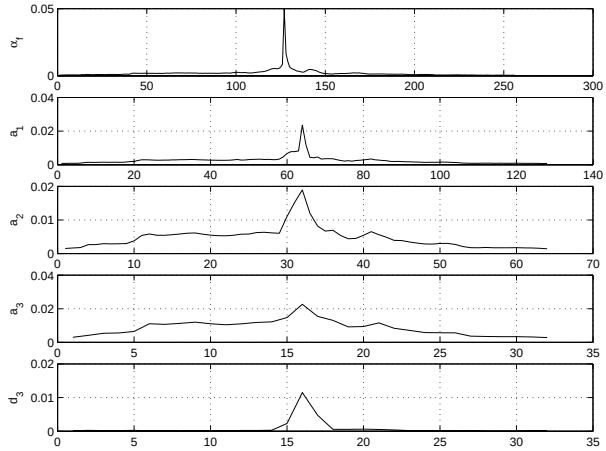


Figure 6.13: The part of the tree of variance of $\alpha_{\mathrm{r}}[n]$, which is interesting in the used method. The figure starts with the original signal. The remaining figure parts are located by denoting a low-pass filtering with h and high-pass filtering with g, meaning that two low-pass filterings followed by one high-pass, are denoted hhg. The second plot is h, the third is hh, the fourth is hhh, and the last one is hhg.

The same method is applied to $\alpha_{\mathrm{r}}[n]$, using the same wavelet, the Haar wavelet. The interesting part of the tree of variance can be seen in Fig. 6.13. The figure starts with the original signal. The second plot in the figure is h, the third is hh, the fourth is hhh, the last is hhg. Notice the large change from hh to hhh, which results in a significant signal in hhg. This signal is useful for fault detection, since it has relatively large signal parts, and does not contain the near zero frequencies, where disturbances are dominating.
The wavelet filters are in the fault used as normal FIR filters, where the wavelet filter coefficients are used as the coefficients in a FIR filter. This means that the data is not processed block by block. Instead, a filtered signal component is computed at each sample.
This means that a useful filter for fault detection in $\alpha_{\mathrm{r}}[n]$ is found, it is two Haar low-pass filters followed by one Haar high-pass filter. The frequency amplitude plot of this filter is illustrated in Fig. 6.14.
Filtering the test data shown in Fig. 6.10 results in a filter output shown in Fig. 6.15.
A zoom of the fault in Fig. 6.15, is shown in Fig. 6.16.
From this figure it is seen that the filtered signal first has a decreasing curve followed by an increasing one. The first part is the response to the beginning of the fault and the second one is the response to the last part of the signal. This means that the detection of the beginning of the fault, can be performed by the absolute filtered signal's first crossing of a threshold. The end is detected by the fourth crossing of this threshold, (when the signal decreased to be lower than the threshold after it has meet its maximum), where

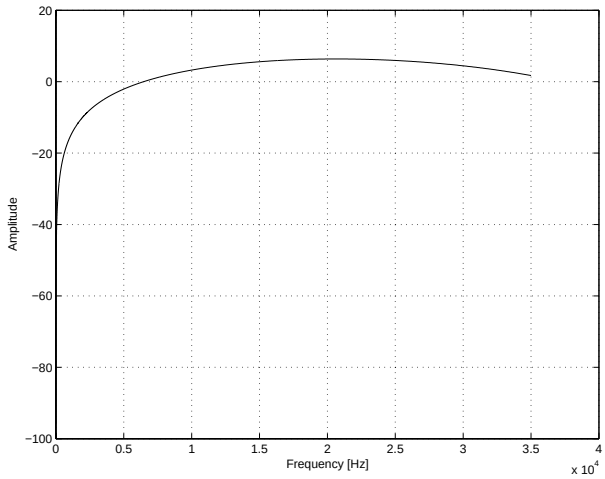


Figure 6.14: The frequency amplitude plot of a filter consisting of two times Haar low-pass filtering followed by one time Haar high-pass filtering.

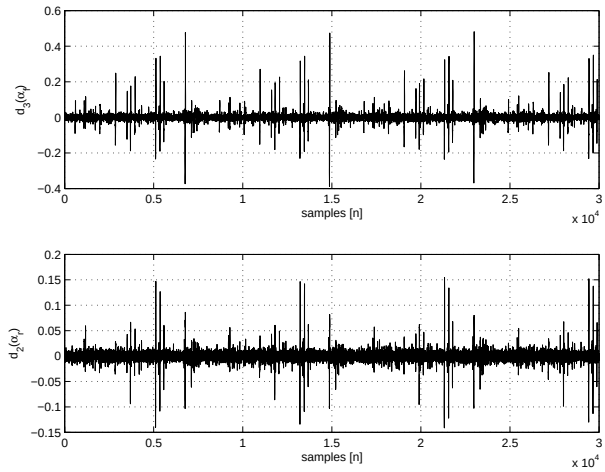


Figure 6.15: The output of using the joint best basis wavelet filters to locate the faults. The upper plot shows filtering output of $e_{\mathrm{f}}[n]$ and the lower plot shows the filtering output of $e_{\mathrm{r}}[n]$.

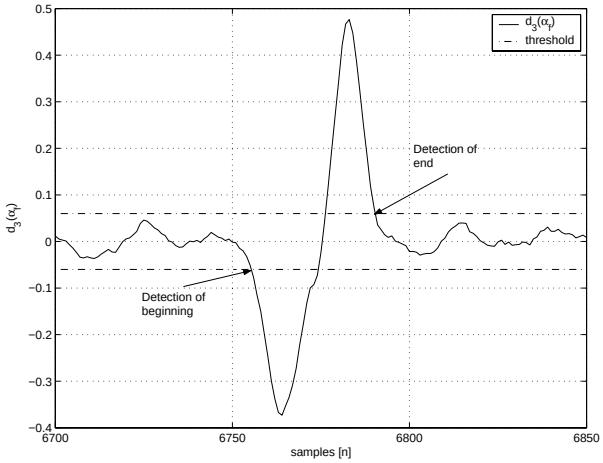


Figure 6.16: A zoom on a scratch in $e_{\mathrm{f}}[n]$ filtered with three Haar low-pass filters followed by one high-pass filter.

α_{f}, α_{r}	Inspection	Normal	WP based
α_{f}	2830-2870	2843-2865	2836-2866
α_{r}	2830-2870	2837-2863	2835-2866
α_{f}	6752-6780	6754-6776	6753-6779
α_{r}	6752-6780	6757-6773	67554-6776
α_{f}	13202-13227	13204-13224	13204-13226
α_{r}	13202-13227	13205-13222	13204-13224
α_{f}	13424-13470	13445-13465	13430-13467
α_{r}	13424-13470	13446-13464	13445-13467

Table 6.2: The time localisation of the four scratch examples, where the wavelet packet based method is compared with time localisation based on visual inspection.

the absolute filtered signal is lower than the threshold. This approach is illustrated by Fig. 6.16.

A threshold is following used on these signals to locate the surface faults in time, see Fig. 6.16, or in other words to detect the faults. To compare with the extended threshold method with the same examples used in Table 6.1, are used in this test see Table 6.2.

Comparing the results of the extended threshold method and the wavelet packet based method, it can be seen that the wavelet packet filter detection is almost as good as the extended threshold methods. However, this method is not preferable since it is not as

good as the extended thresholding method, and it might be sensible to which faults it shall detect. On the other hand the use of other wavelets might improve the time localisation achieved by the use of this method.

6.5 Summary

The topic of this chapter has been time localisation of the surface faults. In the previous chapter, Chapter 5, some interesting time-frequency theories are described, these theories are used in this chapter to perform the time localisation. Section 6.1 has the focus placed on the fact that the surface faults do not vary much from one encounter to the next encounter. In the following three sections, three different methods are used to detect and locate the surface fault based on the decoupled residuals $\alpha[n]$. In Section 6.2 Fang's segmentation algorithm was used, and in Section 6.3 a simple thresholding algorithm was extended in order to cope with the CD-player, and finally in Section 6.4 wavelet packet filters were designed to detect the faults. The conclusion on these three detection and localisation methods is that: Fang's segmentation algorithm is not useful in this application since the achieved filters are too narrow-banded to be used for all surface faults, the extended thresholding method gives a clear improvement of localisation of the surface faults. The wavelet packet filters give an improvement in the localisation of the fault, but not as good as the extended threshold method. However, the use of other wavelet packet filters might improve the achieved results.

Chapter 7

Time-Frequency Based Feature Extraction of Surface Faults

Some work has been performed in handling surface faults. Vidal, see [Vidal Sánchez, 2003], uses robust estimation to estimate the surface faults. However, this estimation was not based on a time-frequency basis. In the Chapter 5 four different time-frequency bases were described. These bases are interesting since they make analysis of the surface faults possible where both time and frequency information are used. This means that the usage of these bases makes it possible to extract some interesting features from the residuals, focus and radial distances which describe the surface faults. Since these features in some way describe the surface faults they can be used to handle the surface faults. These features can be used to remove the effect from for example a scratch on a CD by subtracting an estimate of the fault from the measurements. The feature extract do also provide a classification of the surface faults. In this chapter the feature extraction will be described, starting with a definition of the extracted features, see Section 7.1. The first feature extracted is the fault classification of the surface faults, see Section 7.2, and following by approximating coefficients are computed in Section 7.3.

7.1 Interesting features

In order to use feature extraction to adapt the feature based controller to handle the surface faults, it is needed first of all to identify the usable features for the feature based controller. This feature based controller is in principles an active fault tolerant control scheme is used, meaning that the controllers are accommodated to handle the fault when it occurs. This accommodation is performed based on a fault detection and isolation performed on the residuals. By using a signal processing terminology, detection of the

fault it the same as time localisation of the fault, which was the topic of Chapter 6. The group of all surface faults are a large group and the members of the group vary significantly in time duration and frequency content. In order to adapt the feature based controller to the given fault, it might be a good idea to classify the fault and use this classification to adapt the controller to this given fault class Another important feature for the handling of the faults is approximating coefficients of the faults, which at the same time are poorly approximation of the disturbances etc. The approximation of the surface faults can be used to remove surface fault from the distance measurements, as well as simulations of CD-player playing a CD with a surface fault. This leads to the following features, which will be described in more details together with the extraction of them later in this chapter.

- Classification of the fault
- Approximating coefficients of the fault

7.2 Fault classes and fault classification

In this section the faults will be classified. The set of all surface faults is a large set and the design of one controller handling all faults would in general be a conservative controller. Instead, it would be better to discriminate the detected and time located faults into a number of classes, and use controllers adapted to the given fault group. Some work has been done on automatic classification of surface faults, see [van Helvoirt et al., 2004].
In this work four classes of surface faults has defined based on experimental data. Two of these normally encountered faults classes fingerprints and small scratches, can be merged into one class since fingerprints seen from the control loops appear as a collection of small scratches. Larger scratches have a longer time duration and other frequency contents. This means that the optimal handling of these two classes is not the same. The last class is a class of disturbances like faults, they are caused by other disc faults such as eccentricity, non constant reflection rate of the disc etc. Their frequency contents are in a lower frequency range than the other faults.
As a consequence of the repetitive character of the faults, it is possible to use more time for the feature extraction of the fault, see Section 6.1. In this case more processing time can be used on the discrimination of the fault into the three classes. This also means that the entire fault is available for the discrimination algorithm, and not only a small part of it.
Each fault is inside a window of N samples. In order to simplify the discrimination mentioned above, the faults is transformed into some approximating bases: The Karhunen-Loève basis [Mallat, 1999], the Haar basis [Mallat, 1999], and the mean of class basis. The dimensions of these approximating subspaces are decreased by finding the most Local Discriminating Basis vectors, (LDB), see [Coifman and Saito, 1994] and [Saito, 1994], where the Fisher discriminator is used as cost function, see

[Johnson and Wichern, 2002] and [Flury, 1997]. In this section the 1-10 most discriminating basis vectors are used for the discrimination. These are in addition compared with a set consisting of the set of the means of the vectors in each group. This set of vectors do not span $\mathcal{R}^n$, but is a good comparison for the other discriminating bases.
In this section fault classes are first defined based on experimental data. This data set is divided into a test and a training set. The various bases are shortly described, as well as the algorithm for finding the local discriminating basis vectors. This is followed by a description of the decision rule based on the transform into these LDB vectors. In the end the different discriminating bases are compared based on the test data set.

Fault classes The test data used for the classification and definition of the fault classes are the same data used for training of the wavelet packet filters for time localisation of the faults, see Section 6.4.
This extraction gives two matrices with faults. The faults in $\alpha_{\mathrm{f}}[n]$ are in A_{f} and the faults in $\alpha_{\mathrm{r}}[n]$ are in A_{r}, where each column in the matrices is a fault vector. All faults have by visual inspection been classified into three classes: G_1 which is small scratches, G_2 which is disturbance-like faults, and G_3 which is large scratches.
From each of the classes a training and a test set were formed by randomly taking 80% of the set to be the training set and keeping the remaining part as test sets.

7.2.1 Discriminating algorithm

The faults are in a block of 256 samples in time. In order to discriminate between the different kind of faults, the use of $\mathcal{R}^n,\ n = 256$ is too large. It would instead be a better idea to use some approximating subspaces to reduce the needed order, see [Wickerhauser, 1994a], and subsequently find the most local discriminating basis vectors in this given basis, and e.g. use the q most discriminating ones for the discrimination between the classes, where q is determined by test. The number of the discriminating basis vectors should be low, this means that the search for the optimal value of q is chosen to be in the interval: $[1, 10]$.
The local discriminating subspace is the set of the q most discriminating basis vectors in a given basis, given the classes which shall be discriminated among and a cost function, see [Saito, 1994] and [Coifman and Saito, 1994]. In this section the Fisher discriminator, see [Johnson and Wichern, 2002] and [Flury, 1997], is used as the cost function for finding the most discriminating basis vectors.

7.2.1.1 Fisher discriminator

The Fisher discriminator gives the discriminating power of a number of classes in a given basis, see [Johnson and Wichern, 2002] and [Flury, 1997]. Given an orthogonal basis : $\{x_1, \ldots, x_n\}$, and $S = \{t_m : m = 1, \ldots, M\}$ signals in G_1, and $T = \{t_k : k = 1, \ldots, K\}$ signals in G_2, the discriminating power of the basis vector $\mathbf{x}_i$ between

classes 1 and 2, is defined as:

$$FD(G_1, G_2|\mathbf{x_i}) = \frac{\left|E(< S, \mathbf{x}_i >) - E(< T, \mathbf{x}_i >)\right|^2}{Var(< S, \mathbf{x}_i >) + Var(< T, \mathbf{x}_i >)}, \tag{7.1}$$

and for the basis as a whole:

$$FD(G_1, G_2|\mathbf{x}) = \sum_{i=1\ldots n} FD(G_1, G_2|\mathbf{x_i}). \tag{7.2}$$

A good discriminating basis would have high discriminating power in a few basis vectors and almost nothing in the remaining majority of vectors, and a poorly discriminating basis would have the same discriminating power for all the basis vectors.
In this work the basis is used to discriminate among three classes. This means that the discriminating powers among all the classes for each basis vector is computed:

$$FD(\mathbf{G}|\mathbf{x}_i) = FD(G_1, G_2|\mathbf{x}_i) + FD(G_1, G_3|\mathbf{x}_i) + FD(G_2, G_3|\mathbf{x}_i). \tag{7.3}$$

When all these discriminating powers were computed, the q most discriminating basis vectors were found by choosing the q basis vectors with the highest discriminating powers.

7.2.2 The test discrimination bases

Three different bases are tried in this work: The Karhunen-Loève basis, a generalised Haar basis and the basis of means of classes.

7.2.2.1 Karhunen-Loève basis

The potentials of this basis are explored since it is the best approximating linear basis for a given training set. It is used to reduce the dimension of the subspace, for which the best discriminating basis vectors are found. The Karhunen-Loève basis is described in more details in Section 5.5. These eigenvectors/ Karhunen-Loève basis vector of the data set, are eigen faults, (the notation refers to Wickerhauser's notation of eigenfaces in [Wickerhauser, 1991]). The approximating property of this basis is illustrated in Fig. 7.1 where a time series of $\alpha_{\mathrm{f}}[n]$ containing a fault, is approximated with one and four Karhunen-Loève basis vectors. From this it is clear that just a few Karhunen-Loève basis vectors give a good approximation of the original signal.

7.2.2.2 Haar basis

Wavelet bases in general and the Haar basis specifically are much more simple (faster in computations) than the Karhunen-Loève basis, but on the other hand not as good approximating basis. The following generalised Haar bases potentials as a basis are

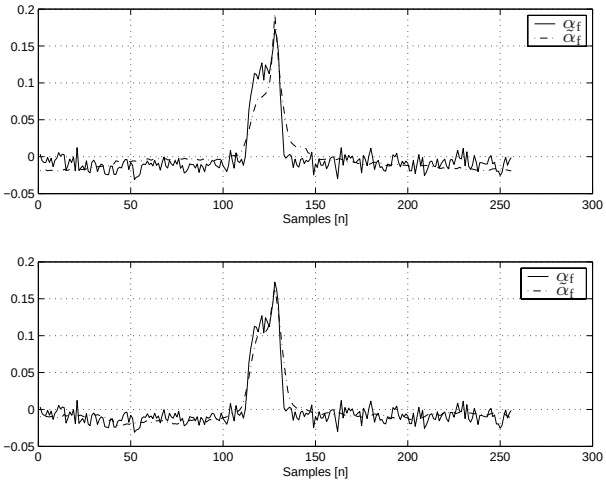


Figure 7.1: Illustration of the Karhunen-Loève approximation of α_{f} which contains a typical scratch. The approximation is denoted with $\tilde{\alpha}_{\mathrm{f}}$. The first approximation is based on the most approximating coefficient, the second approximation is based on the four most approximating coefficients.

explored, since it is a simple basis. It is a generalised Haar basis since some other properties are needed than those given by the normal Haar basis. No reconstruction is needed based on this basis. However, du to the Fisher discriminator it needs to be almost orthogonal. It is still a normalised basis. In addition it has to be time invariant, which is easily obtain, since there is no requirement of reconstruction. The basis vectors, w_i, have the length of $2^8 = 256$. For the vector number N it is formed as follows. The first N elements take the value 1, the next N elements take the value -1, and the remaining ones take the value zero. The vector is next normalised by multiplication of the factor 2^{-n}. The first 8 vectors are defined in this way. The last vector is the maximum value of the signal which shall be transformed. The basis vectors $w_1, \cdots, w_8$, are illustrated in Fig. 7.2. These basis vectors are all orthogonal.
Since the faults cannot be assured to be centred in the data set, this transform has to be time invariant. Since this transform is only used for analysis, it can be handled simple by computing the coordinates, c_j, by:

$$c_j = \max(|s_{\mathrm{m}} * w_j|). \tag{7.4}$$

s_{m} denotes the sampled signal, and w_j the j'th basis vector. The bases' orthogonality property is lost in this time invariance handling. However, it is close to be orthogonal. The orthogonality is only lost if the maximum of the convolution relates to a time shift in the basis vector which makes the two basis vectors changing sign at the same sample.

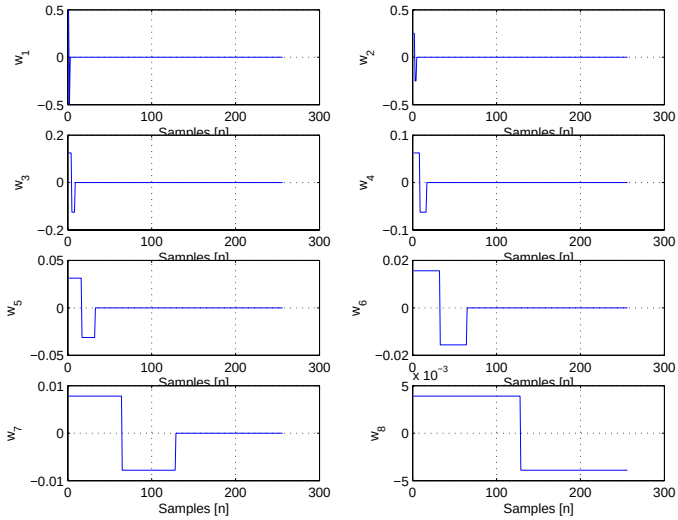


Figure 7.2: The 8 generalised Haar basis vectors given the equations in Subsection 7.2.2.2.

7.2.2.3 Mean of the classes

Normally a good way to discriminate among signals is to convolute the signal with the signals them self, and the convolution giving the highest result is the convolution with the signal itself. But if the signal is not perfectly known or one which is a discrimination of classes containing more than one signal, this method is not so good anymore. However, these arguments indicate the usability of a discriminating basis consisting of the mean of the classes. This set is not orthogonal nor does it span $\mathcal{R}^n$, but it is a good set of vectors for the other bases to compare with, since it normally has good discriminating properties.

7.2.3 Finding the discriminating basis vectors

The Fisher discriminating power function is following used to find the four most discriminating vectors in each basis. These four most discriminating basis vectors are following used for discrimination.

7.2.3.1 The algorithm

For each class, the centre of the class in the given discriminating basis, is computed based on the training set. This gives three vectors: C_1, C_2, C_3.
The discriminating algorithm is: Find the class which has a centre with the smallest

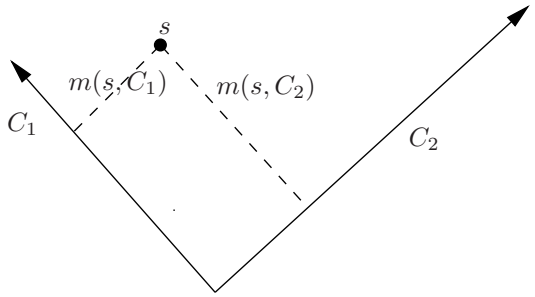


Figure 7.3: Illustration of the decision rule of the discriminator. s is the fault in question. C_1 and C_2 are the centres of two classes. $m(s, C_1)$ and $m(s, C_2)$ are the measures of the distances between the fault in question and the respective group centres. Since s is closest to C_1, this means that $m(s, C_1) < m(s, C_2)$. The decision is that the fault in this example is a member of class 1, since it is closest to centre of class 1.

distance to the sample, given a metric. A successful measure has been found to be the geometric distance between the sample s and the vectors defined as the coordinates of the centre:

$$m(s, C_i) = \left\| s - < C_i, s > C_i \right\|_2 . \tag{7.5}$$

The decision rule and measure are illustrated in Fig. 7.3. In this illustration the fault in question, s, is closest to the centre of class 1, and the fault is as consequence a member of class 1. To illustrate the algorithm's success, an array is defined, where the rows indicate which class the fault is contained by, and the column which classes they are detected as.

7.2.4 Results of fault classification

The discriminator's results are computed for the 1 to 10 most discriminating basis vectors. For 1 to 2 basis vectors the results are not impressive. For 3 to 6 basis vectors, the improvement of the results is clear. The results of the discriminator based on the Karhunen-Loève basis, the Haar basis and the mean of class sets are given in Tables 7.1, 7.2 and 7.3. Before choosing the best discriminator from these results in Tables 7.1-7.3, it is necessary to define some requirements to the discriminator. The most important issue is to classify G_3-faults as being G_3-faults, since the controllers are maybe forced into severe problems if these faults are not classified correct. The second most important thing is to have as high correct classification of G_1 faults. G_2-faults can be detected in another way. Another important issue is to limit the required computations. This means that a low number of basis vectors is better than a high number of basis vectors. It also means that the Haar basis has a disadvantage in the way in which the basis transformation is done. It is made time invariant by finding the max of the auto correlation of the

Karhunen-Loève basis α_{f}				Karhunen-Loève basis α_{r}			
3 Basis vectors				3 Basis vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	85.7%	25.0%	0.0%	G_1	83.8%	0.0%	0.0%
G_2	9.1%	50.0%	0.0%	G_2	6.8%	0.0%	16.7%
G_3	5.2%	25.0%	100%	G_3	9.5%	100%	83.3%
4 Basis vectors				4 Basis vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	87.0%	25.0%	0.0%	G_1	85.1%	0.0%	0.0%
G_2	10.4%	50.0%	0.0%	G_2	8.1%	0.0%	0.0%
G_3	2.6%	25.0%	100%	G_3	6.8%	100%	100%
5 Basis vectors				5 Basis vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	87.0%	25.0%	0.0%	G_1	83.8%	0.0%	0.0%
G_2	10.4%	50.0%	0.0%	G_2	8.4%	0.0%	16.7%
G_3	2.6%	25.0%	100%	G_3	9.5%	100%	83.3%
6 Basis vectors				6 Basis vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	89.6%	50.0%	0.0%	G_1	83.8%	0.0%	0.0%
G_2	9.1%	25.0%	0.0%	G_2	2.7%	100%	0.0%
G_3	1.3%	25.0%	100%	G_3	13.5%	0.0%	100%

Table 7.1: The results of the discriminator based on the 3 to 6 most discriminating Karhunen-Loève basis vectors are shown in this table. The left half part of the table is based on the residual $\alpha_{\mathrm{f}}[n]$, and the other half part on the residual $\alpha_{\mathrm{r}}[n]$. These parts are again separated into arrays with the results of each number of basis vectors. The horizontal $G_1,\ G_2,\ G_3$ are the class in which the test data are contained and the vertical ones are the classes they are classified as being in. This can be illustrated by an example. Take α_{f} with 3 basis vectors. 85.7% G_1-faults were classified as being G_1-faults, 9.1% were classified as being G_2-fault and the remaining 5.2% were classified as being G_3-faults.

basis vectors and the residuals, where the other two basis transformations are done by convoluting the basis vectors with the residuals. This means that a Haar based approach has to perform significantly better than the other bases to be chosen, as the best one, and in addition it is time invariant.

Inspection of the three tables with the results show that the best performance is achieved by using the Haar transform with 5 basis vectors on $\alpha_{\mathrm{r}}[n]$, where G_1-faults were classified correct with 94.6% success, and the two other classes were correctly classified with 100% success. The mean-of-class basis achieves the best performance for the $\alpha_{\mathrm{r}}[n]$ residual, with 82.4% success for G_1-faults and 100% success for the two others. The Karhunen-Loève basis performs best at the $\alpha_{\mathrm{f}}[n]$ residual. The Karhunen-Loève basis does not improve its performance much from 3 to 6 basis vectors. It is interesting to compare the result of order 3 and 4 discriminator based on the Karhunen-Loève basis,

Haar basis α_{f}				Haar basis α_{r}			
3 Basis vectors				3 Basis vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	49.4%	50.0%	0.0%	G_1	75.5%	0.0%	0.0%
G_2	19.5%	25.0%	0.0%	G_2	4.1%	100%	66.7%
G_3	31.2%	25.0%	100%	G_3	20.3%	0.0%	33.3%
4 Basis vectors				4 Basis vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	35.1%	25.0%	0.0%	G_1	79.7%	0.0%	0.0%
G_2	42.9%	50.0%	0.0%	G_2	16.2%	0.0%	16.7%
G_3	22.1%	25.0%	100%	G_3	4.1%	100%	83.3%
5 Basis vectors				5 Basis vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	54.5%	25.0%	0.0%	G_1	94.6%	0.0%	0.0%
G_2	31.1%	50.0%	0.0%	G_2	1.4%	100%	0.0%
G_3	14.3%	25.0%	100%	G_3	4.1%	0.0%	100%
Basis order 6				Basis order 6			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	84.4%	25.0%	0.0%	G_1	77.0%	0.0%	0.0%
G_2	10.4%	50.0%	0.0%	G_2	17.6%	100%	0.0%
G_3	5.2%	25.0%	100%	G_3	5.4%	0.0%	100%

Table 7.2: The results of the discriminator based on 3 to 6 most discriminating Haar basis vectors are shown in this table. The left half part of the table is based on the residual $\alpha_{\mathrm{f}}[n]$, and the other half part on the residual $\alpha_{\mathrm{r}}[n]$. These parts are again separated into arrays with the results of each number of basis vectors. The horizontal $G_1,\ G_2,\ G_3$ were the class in which the test data were contained and the vertical ones are the classes they were classified as being in. This is the same principle as in Table 7.1.

Mean set α_{f}				Mean set α_{r}			
3 vectors				3 vectors			
	G_1	G_2	G_3		G_1	G_2	G_3
G_1	79.2%	0.0%	50.0%	G_1	82.4%	0.0%	0.0%
G_2	13.0%	75.0%	25.0%	G_2	17.6%	100%	0.0%
G_3	19.5%	0.0%	25.0%	G_3	0.0%	0.0%	100%

Table 7.3: The results of the discriminator based on the mean of class set of the order 3, this set has only 3 vectors, shown in this table. The left half part of the table is based on the residual $\alpha_{\mathrm{f}}[n]$, and the other half part on the residual $\alpha_{\mathrm{r}}[n]$. The horizontal $G_1,\ G_2,\ G_3$ were the class which the test data were contained in and the vertical ones were the classes they were classified as being in. This is the same principle as in Table 7.1.

with mean-of-class basis based discriminations. The Karhunen-Loève based discriminator achieves 50% success for G_2-faults for both 3 and 4 basis vectors, and 100% for G_3-faults for both the 3 and 4 basis vectors. The G_1 success rate is 85.7% if 3 Karhunen-Loève basis vectors are used and 87% if 4 Karhunen-Loève basis vectors are used.
Even though the Haar basis based discriminator performs the best, it is presumably not preferable due to its high demands of computations due to the time invariant property. However, this property can be important if the faults in question are not symmetrical placed in the fault vectors. Comparing the Karhunen-Loève basis and the mean of class basis, they have the same good performance regarding discriminating G_3-faults. The Karhunen-Loève basis has 3 or 5 % point better success rate of G_1-faults, but it does not perform as well at G_2. However, G_2 can be detected in another way. This means that if the number of computations is not a large problem, the projection of $\alpha_{\mathrm{r}}[n]$ on the 5 Haar basis vectors is the best discriminator. If the number of computations is of concern one should chose the projection of $\alpha_{\mathrm{f}}[n]$ on the 3 most discriminating Karhunen-Loève basis vectors.

7.2.5 Summary of fault classes and fault classification

Based on the test data from real world challenging CDs, three fault classes are defined. A discriminator, is designed to discriminate between these classes. This discriminator is found based on the local discriminating basis of some approximating bases: Karhunen-Loève, Haar etc. After the basis transformation, the discriminator finds the class which the given fault is closest to in the given basis. The Karhunen-Loève basis based detection has rates higher than 85.7% for the important short faults, G_1, and large faults, G_3. The much more computationally demanding Haar basis based discriminator has success rates higher than 94.6% for all the three fault classes. In addition to these high success rates the Haar basis based discriminator is time invariant, which is an important property if it cannot ensured that the faults are symmetrically placed in the middle of the data block. Another possibility is to combine the different bases to achieve a better performance.

7.3 Approximation of surface faults with Karhunen-Loève base

The surface faults are parameterised in Chapter 3, with:

$$\mathbf{s}_{\mathrm{m}}[n] = \begin{bmatrix} \beta_{\mathrm{f}}[n] \cdot \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \beta_{\mathrm{r}}[n] \cdot \mathbf{I} \end{bmatrix} \cdot \mathbf{f}\left(e_{\mathrm{f}}[n] + \check{e}_{\mathrm{f}}[n], e_{\mathrm{r}}[n] + \check{e}_{\mathrm{r}}[n]\right). \tag{7.6}$$

The Karhunen-Loève base has previously been used to approximate the residuals $\alpha_{\mathrm{f}}[n]$, see Section 6. It is obvious to use it to approximate the $\check{e}_{\mathrm{f}}[n]$ and $\check{e}_{\mathrm{r}}[n]$. Based on the

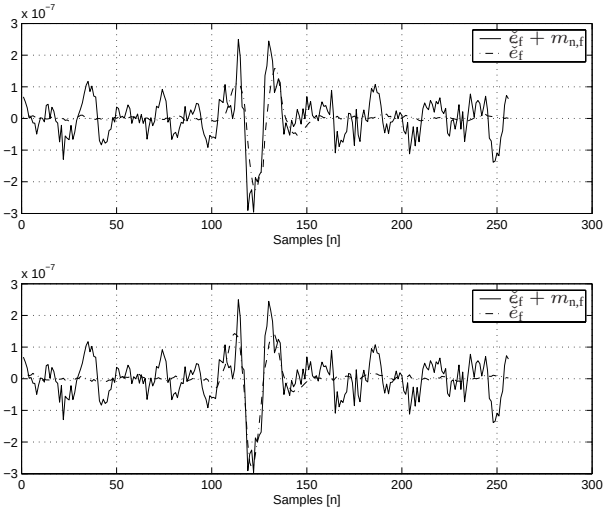


Figure 7.4: Illustration of the Karhunen-Loève approximation of $\check{e}_{\mathrm{f}}+m_{\mathrm{n,f}}$ which contains a typical scratch. The approximation is denoted with $\tilde{\check{e}}_{\mathrm{f}}$. The first approximation is based on the most approximating coefficient, the second approximation is based on the four most approximating coefficients.

same data as used to compute the Karhunen-Loève bases of the residuals, the Karhunen-Loève base of $\check{e}_{\mathrm{f}}[n]$ can be computed. In Fig. 6.16 these are approximated by the use of respective one and four Karhunen-Loève bases vectors. From it can be seen that the four Karhunen-Loève bases vectors approximates the main trends in the signal but the noises. This means that these four Karhunen-Loève bases vectors approximates $\check{e}_{\mathrm{f}}[n]$ an not $m_{\mathrm{n,f}}$. I.e. the fault component in distance signal is separated from the measured signal. Due to variety of the fault classes an approximating basis is computed for each of the three classes of faults.

7.4 Summary

The topic of this chapter has been feature extraction from the surface faults. These features are extracted with the purpose to remove or at least partly remove the influence of the surface faults. In Chapter 5 some relevant time-frequency theory was described. This theory is in this chapter to perform the feature extraction. In Section 7.1 the extracted features are described. These are: class and classification of the fault, and approximation of the fault. In Section 7.2 the classification of the fault is described. The approximation of the fault is described in Section 7.3.

Chapter 8

A Simulation Model

Previous designs of fault tolerant controllers of CD-players have been based on trial and error on real test systems since no simulation models including the surface faults have been available. The CD-player servos consist roughly of two parts: an electro-magnetic-mechanical part - the actuator, and the electro-optical part - the sensor. Another important part of the system is the surface faults on the CDs, if one would like to handle those. These parts are modelled in Chapter 3.

In [Odgaard and Wickerhauser, 2003] some work is done on finding a good approximating basis of the surface faults. It is done by finding the Karhunen-Loève basis, see Chapters 5 and 6, on a set of experimental data with faults.

By using a Karhunen-Loève basis, based on the experimental data, the faults can be approximated with a few coefficients. This can be used to simulate faults, by computing the required coefficients by random numbers with the same mean and variance as the training data did have.

In this chapter the model of the CD-player and surface faults are summarised, see Section 8.1. The simulation model is developed based on a Karhunen-Loève approximation of surface faults, and the mean and variance of the experimental data. In addition disturbances and measurement noises are added at the relevant places in the system, see Section 8.2. This chapter does not deal with fault tolerant control, but with a simulation model of CD-player. This model simulates both the focus and radial loop, this means that simulations require both a focus and a radial controller. In order to show this model in practice two simulations are performed in closed loop with two different pairs of PID-controllers with and without a surface fault. The first simulated pair of controllers have a high bandwidth which makes it good for handling disturbances. The second simulated pair of controllers have a lower bandwidth which makes it better suited for handling surface faults, see Section 8.3.

8.1 Model

The idea of the discrete time model is to simulate focus and radial servo loops in CD-players with relevant disturbances as well as surface faults represented by models based on experimental data. This model of the focus and radial servos in the CD-player is derived in Chapter 3. The simulation model includes some extra parts compared with the model derived in Chapter 3. The overall structure of this simulation model is illustrated in Fig. 8.1. The controllers' jobs are to keep the OPU focused and radially tracked on the disc information track. The actuator used to move the OPU in focus and radial directions is a 2-axis linear electro magnetic actuator. This Electro Mechanical System (EMS) is used to set the OPUs absolute position, $\mathbf{x}[n] = \begin{bmatrix} x_{\mathrm{f}}[n] \\ x_{\mathrm{r}}[n] \end{bmatrix}$. The track position is given by the position reference $\mathbf{x}_{\mathrm{ref}}[n] = \begin{bmatrix} x_{\mathrm{ref,f}}[n] \\ x_{\mathrm{ref,r}}[n] \end{bmatrix}$. These references are due to eccentricity and skewness of the disc, since these dominates the inherited reference given by the spiral shaped track. I.e. the vector of distance signals can be defined as: $\mathbf{e}[n] = \mathbf{x}_{\mathrm{ref}}[n] - \mathbf{x}[n]$. The Optical System, OS, is used to retrieve data from the disc and to generate four detector signals, $\mathbf{s}[n] = \left[D_1[n],\ D_2[n],\ S_1[n],\ S_2[n]\right]^T$, which are used to approximate the focus and radial errors. This approximation is done in the Signal Converter, SC. The SC computes normalised approximations of the error signals, $\mathbf{e}_{\mathrm{n}}[n]$. In Chapter 3, the entire CD-player model is derived as follows

$$\mathbf{s}_{\mathrm{m}}[n] = \begin{bmatrix} \beta_{\mathrm{f}}[n] \cdot \mathbf{I}_{2\times 2} & \mathbf{0} \\ \mathbf{0} & \beta_{\mathrm{r}}[n] \cdot \mathbf{I}_{2\times 2} \end{bmatrix} \cdot \mathbf{f}\left(e_{\mathrm{f}}[n] + \check{e}_{\mathrm{f}}[n], e_{\mathrm{r}}[n] + \check{e}_{\mathrm{r}}[n]\right), \tag{8.1}$$

where $\mathbf{f}(\cdot)$ is the model of the optical detectors, see (3.13)-(3.51). Focus and radial distances $e_{\mathrm{f}}[n]$ and $e_{\mathrm{r}}[n]$ are modelled in (3.11) and (3.12). $\beta_{\mathrm{f}}[n], \beta_{\mathrm{r}}[n]$ are two fault model parameters. This fault model can be explained, but unfortunately a fault affects the detector signal in another more problematic way. Surface faults also introduce faulty focus and radial errors, $\check{\mathbf{e}}[n]$, which are added to $\mathbf{e}[n]$. The task of the controller is to keep the OPU focused and radially tracked. It is done by feeding the EMS with control signals $\mathbf{u}[n]$ calculated from past and present values of $\mathbf{e}_{\mathrm{n}}[n]$. In order to simulate external disturbances like shocks and internal not model selfpollutions, a vector of disturbance signals, $\mathbf{d}[n]$ is added to $\mathbf{u}[n]$ before they are fed to the EMS.

8.2 Surface fault synthesiser

The surface faults are parameterised in (8.1). The next step is to synthesise the values of these parameters during a fault. One approach would be to take measured data from a real CD-player playing a CD with a surface fault and use these measured data in the simulation representing a surface fault. This approach has a severe drawback, it limits the number of possible faults in simulation to those measured. In this chapter another

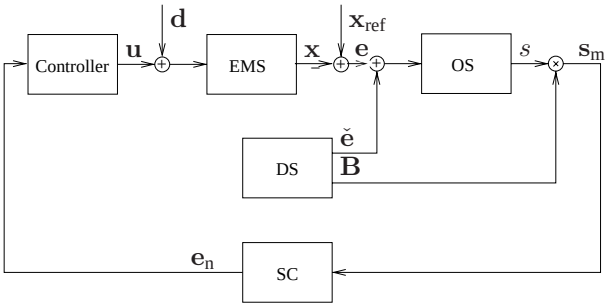


Figure 8.1: Illustration of the structure of the CD-player simulations model. The sub-models in the models are as follows. EMS, denotes the Electro Mechanical system, OS is the Optical System of the OPU. DS denotes the Fault Synthesiser, and simulates surface faults. SC denotes the Signal Converter which computes normalised focus and radial differences which are approximations of focus and radial distances. In addition focus and radial disturbance, $\mathbf{d}[n]$, and an unknown references, $\mathbf{e}_{\text{ref}}[n]$, are added to the system.

approach is presented, which is based on Karhunen- Loève approximations of the faults, see Chapter 6.
Blocks of 256 samples are extracted from experimental data, where each block contains at least one fault. The experimental data represent fingerprints, short and long scratches etc. In Section 3.1 the experimental setup used to measure the data is described. By inspection the experimental data are separated into three classes in Section 7.2: short and long scratches and disturbance like faults.
The next step in this approach is to compute the best approximating basis of this data set - the Karhunen-Loève basis, see Chapter 5 and [Mallat, 1999]. The approximating potentials are illustrated in Fig. 7.2. This illustrates a scratch approximated with the most approximating coefficient and the four most approximating coeffiecients respectively. Note the approximation with only one coefficient is a good approximation of the scratch.
Denote a matrix, where the column vectors are vectors of zero mean data, $\mathbf{S}$. The Karhunen-Loève basis, see Section 5.5, $\mathbf{K}$, of $\mathbf{S}$ is the set of eigenvectors of the autocorrelation of $\mathbf{S}$:

$$\mathbf{K} = \{\text{eigenvectors}\left(\mathbf{S}\mathbf{S}^T\right)\}. \tag{8.2}$$

The Karhunen-Loève basis of all the measured $\beta_{\text{f}}[n]$ is denoted $\mathbf{K}_{\beta_{\text{f}}}$. $\mathbf{K}_{\beta_{\text{r}}}$ denotes the Karhunen-Loève basis of measured $\beta_{\text{r}}[n]$. $\mathbf{K}_{\check{e}_{\text{f}}}$, $\mathbf{K}_{\check{e}_{\text{r}}}$ denote the Karhunen-Loève basis of $\check{e}_{\text{f}}[n]$ and $\check{e}_{\text{r}}[n]$ computed based on measurements, which also contain measurement noises etc.
Assume that each of the measured faults consists of a signal part and a noise part. Using the Karhunen-Loève basis the signal part can be approximated with a few coordinates

in that basis, (the first k coefficients). The remaining noise part could be approximated with the remaining, $N - k$ coefficients. Next compute the mean m_i and variance σ_i of each coordinate. Assume that each Karhunen-Loève coordinate can be simulated as independently distributed normal random variables with σ_i and m_i, where these statistics are found based on the Karhunen-Loève coordinates. The signal part can then be synthesised as:

$$\mathbf{S}_{\text{syn}} = \sum_{i=1}^{k} \mathbf{K}_{\text{syn},i}\mathbf{c}_{k,i}, \tag{8.3}$$

and the noise as

$$\mathbf{S}_{\text{syn,N}} = \sum_{i=k+1}^{N} \mathbf{K}_{\text{syn},i}\mathbf{c}_{k,i}, \tag{8.4}$$

where

$$\mathbf{K}_{\text{syn}} \in \{\mathbf{K}_{\beta_\text{f}}, \mathbf{K}_{\beta_\text{r}}, \mathbf{K}_{\check{e}_\text{f}}, \mathbf{K}_{\check{e}_\text{r}}\}, \tag{8.5}$$

$$\mathbf{c}_{k,i} \in \mathcal{N}_i(m_i, \sigma_i), \tag{8.6}$$

and $\mathbf{S}_{\text{syn}}$ denote the synthesised $\beta_\text{f}, \beta_\text{r}, \check{e}_\text{f}, \check{e}_\text{r}$, and $\mathbf{S}_{\text{syn,N}}$ the noise part of these. Since $k \ll n$ the signal synthesis is cheap in computations, but the noise part is expensive. Instead the noise signal can be computed as

$$\mathbf{S}_{\text{syn,N}} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix} \tag{8.7}$$

, where $y_i \in \mathcal{N}(0, \frac{\sigma}{\sqrt{N}})$ is independent, σ is equal $\sqrt{\sigma_{k+1}^2 \ldots \sigma_N^2}$. The coordinates in the noise vector are independent, identically distributed normal random variables, since the following is well known, If T:$\mathcal{R}^N \rightarrow \mathcal{R}^N$ is orthonormal and y is defined in (8.7), then the coordinates of Ty are independent, identically distributed normal random variables signal with same mean, 0, and variance, $\frac{\sigma}{\sqrt{N}}$.

It is now possible to synthesise a fault. The statistics used for the fault construction is based on all faults in a fault class. However, a fault does not vary much from one encounter to the next encounter, see Section 6.1. This fact has to be taking into account when the next fault encounter is simulated, see Section 5. σ_i is computed as the mean of the variances of coefficients of sequently following encounters of faults. The mean used in the construction is the simulated coefficient value from the last fault encounter.

This fault synthesiser constructs the faults in blocks of 256 meaning that it is required to compute a long block of fault parameters before the simulation of the CD-player, and then during the simulation use these computed values.

8.2.1 Disturbances and references

The CD-player's unknown reference is mainly due to eccentricity and skewness. It is computed as a sine with the period being the rotation period of the given rotation. These periods increase with one sample each third rotation, due to the linearly increased rotation length measured in number of samples, due to the spiral shaped information track on the disc. The amplitude of the reference can freely be chosen, (with the player requirements stated in [Stan, 1998]), and in this simulation both are chosen to be 30 μm. The disturbances are white noise added to the control signals, where the variance 0.01 is chosen based on experience.

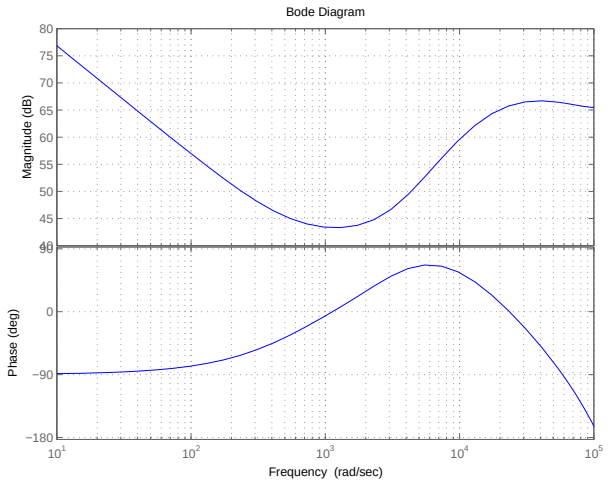


Figure 8.2: Bode plot of the nominal focus controller.

8.2.2 Signal converter

The Signal Converter is described in Section 3.5, where it is described that focus and radial distances are approximated by the use of the normalised difference between D_1 and D_2 for focus and S_1 and S_2 for radial. This means

$$\mathbf{e}_{\mathrm{n}} = \begin{bmatrix} \frac{D_1 - D_2}{D_1 + D_2} \\ \frac{S_1 - S_2}{S_1 + S_2} \end{bmatrix} \tag{8.8}$$

8.2.3 Controller

The intended use of this model is to simulate the performance of the designed controllers before they are implemented on a real system. As a consequence it is possible to try any controller on the simulation model. In the simulations in this chapter two pairs of simple PID-controller are used, the pair consist of a focus and radial controller. The first pair are the typical PID-controllers described in [Stan, 1998] with a gain adapted to the given CD-players optical gains, see Fig. 8.2. The second pair have a lower bandwidth in order to handle surface faults better. A bode plot of the focus controller is shown in Fig. 8.3. The four controllers can be seen in Appendix D. This simulation model gives a

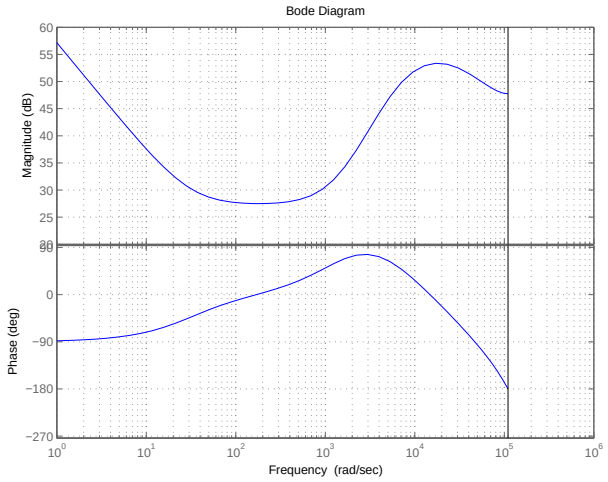


Figure 8.3: Bode plot of a focus controller with lower bandwidth.

method to validate controller performances regarding handling of faults. This given by the following definition.

8.1 Definition *Define* $\mathbf{e}[n]$ *in the fault free case as* $\mathbf{e}_1$ *and in the case of a fault as* $\mathbf{e}_2$. *Then the fault handling performance,* $\mathcal{DP}$*, can be defined as:*

$$\mathcal{DP} = \frac{\left\|\mathbf{e}_1 - \mathbf{e}_2\right\|_2}{\left\|\check{\mathbf{e}}\right\|_2} \tag{8.9}$$

.

This index computes the controllers sensitivity towards a simulated fault, since $\mathbf{e}_1 - \mathbf{e}_2$ is equal the CD-player's response to the fault. Remark also that all the signals in 8.9 are finite signals.

8.3 Simulations of a CD-player playing a CD with surface faults

The simulations are performed in closed loop, where the two simulated controllers control the CD-player model, with the objective to suppress the disturbances and the unknown reference. The first simulations of the two controllers are simulations without any surface faults. These fault free simulation are important in order to validate the given controllers capability to handle a fault, since any differences from these are due to the fault, and not the controllers ability to suppress disturbances. These signals are as consequence used in Def. 8.1 where these errors are denoted $\mathbf{e}_1$.
The controllers react to a sum of the real distances and the faulty distance components. To illustrate how the controller forces the system out of focus and radial tracking during a fault, the simulated real distances are illustrated and not the distances used for controlling the system, since they contain the faulty component too. The simulation output of the normal PID-controller can be seen in Fig. 8.4 which is a zoom on a fault. The second subfigure in Fig. 8.4, is a zoom on the fault, $\check{e}_{\mathrm{f}}[n]$. This means that the controller reacts on the focus error component due to the fault, $\check{e}_{\mathrm{f}}[n]$, which is undesired. The controller with a lower bandwidth is illustrated as well in Fig. 8.4. This figure shows that the low bandwidth controller handles the fault better than the normal one. The upper plot in Fig. 8.4 do also validate the simulation model, since this close loop output of the nominal controller, is similar to the measured focus distance during a surface fault where the focus distance is controlled by the nominal controller.
The two controllers can be compared by using Def. 8.1. The computed indexes can be seen in Table 8.1, in addition the nominal performances, $\mathcal{NP}$, are also computed defined as:

$$\mathcal{NP} = \frac{\text{rms}\left(\mathbf{e}_1\right)}{\text{rms}\left(\mathbf{x}_{\text{ref}}\right) + \text{rms}\left(\mathbf{d}\right)} \tag{8.10}$$

The two-norm of the error signal can be use to compare the nominal performance of the two controllers since the same reference, disturbances and faults are applied to them in the simulations. The performance index consists of two elements the first is the focus index and the second element is the radial index. Based on these results it can be seen that the nominal controller is better for handling disturbances, but on the other hand the low bandwidth PID-controller is better for handling the surface faults. This supports the known controller specification conflict between handling faults and disturbances. One should notice that the simulated fault in this situation is a small fault, which controllers normally survives, but in a non-optimal way. Handling larger faults cause larger problems for the controller, meaning that a fault tolerant approach is relevant in order not to react to surface faults, where surface faults are viewed as being sensor faults. The use of this simulation model has a potential to make the design of controllers handling surface fault faster, since the use of simulations can eliminate some of the practical experiments in the controller design process.
The simulations are implemented in MATLAB™. The used MATLAB™script files and

	Normal PID Controller	Low bandwidth Controller
$\mathcal{DP}$	$\begin{bmatrix} 1.2785 \\ 0.9219 \end{bmatrix}$	$\begin{bmatrix} 0.5574 \\ 0.1845 \end{bmatrix}$
$\mathcal{NP}$	$\begin{bmatrix} 3.7368 \cdot 10^{-4} \\ 2.7317 \cdot 10^{-4} \end{bmatrix}$	$\begin{bmatrix} 8.9161 \cdot 10^{-4} \\ 12.3812 \cdot 10^{-4} \end{bmatrix}$

Table 8.1: Table of the fault handling performance $\mathcal{DP}$ and the nominal performance $\mathcal{NP}$, for both the normal PID-controller and the low bandwidth PID-controller. The performance indices are vectors with the first element being the focus index and the second the radial index.

data files can be seen at [Odgaard et al., 2003d].

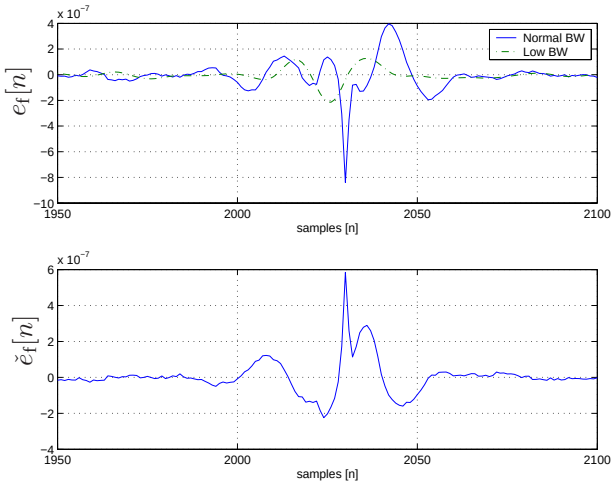


Figure 8.4: The upper plot is a zoom on a surface fault in $e_{\mathrm{f}}[n]$, for both controllers. Notice the low bandwidth controller reacts less on the fault. The lower plot is of $\check{e}_{\mathrm{f}}[n]$, for the same fault. $e_{\mathrm{f}}[n]$ is strongly correlated to $\check{e}_{\mathrm{f}}[n]$ this means that the controller react on the faulty error signal. I.e. the controller handles this fault in an undesirable way.

8.4 Summary

In this chapter a closed loop simulation model of a CD-player playing a CD with surface faults is presented. The use of this simulation model has a potential to make the design of controllers handling surface faults faster, since the use of simulations can eliminate some of the practical experiments. In Section 8.1 the model of the CD-player is summarised.

In Section 8.2 a surface fault synthesiser is derived. The simulations show that the focus and radial servo, in a typical CD-player, react on the fault in a non-desired way. In addition a controller with a lower bandwidth is simulated. A performance index of the controllers handling fault is also defined and used to compare the two simulated controllers. This comparison validates a known fact that handling disturbances in CD-players requires a high bandwidth and handling surface faults on CDs requires a low bandwidth, see Section 8.3.

Chapter 9

Feature Based Control of Compact Disc Players

In this chapter the core of this feature based control scheme of the CD-player will be described and designed. This algorithm is based on the work presented in the previous chapters. The idea in short is, that the residuals computed in Chapter 4 are used to detect and locate the surface faults, by the use of the algorithms presented in Chapter 6. The feature extraction presented in Chapter 7 gives a classification of the fault as well as approximating coefficients in a given basis representing the class of the fault. The feature based control algorithm is presented in the following sections and in [Odgaard et al., 2004c] and [Odgaard et al., 2004b]. In Section 9.1 the feature based control scheme is introduced and described in general terms and related to the fault tolerant control method. This leads to fault accommodation part of the feature based control algorithm which is described in Section 9.2. Some stability and performance issues are discussed in Section 9.3, where it is proven that the method is stable. In Section 9.4 the simulation model derived in Chapter 8 is used to simulate the feature based control scheme handling a CD with a scratch. These simulations are followed by experimental results presented in Section 9.5.

9.1 Feature based control versus fault tolerant control

An overview of the feature based control scheme was presented in Section 2.9. The feature based control strategy is illustrated in Fig. 2.12, from which it can be seen that the feature based control strategy consists of: residual and distance estimation, fault detector/fault locator, feature extraction, fault accommodation. The residual and distance estimation estimates the decoupled residuals and focus and radial distances. The fault detector/locator detects and locates the fault in time. The time localisation gives infor-

mation of when to extract features and when to accommodate the fault. The feature extraction feeds the fault accommodation with the class of the fault and approximating coefficients in an approximating basis of the given class of the fault. This feature based control strategy is strongly related to fault tolerant control scheme. In order to explain this relation, the fault tolerant control method is presented.

Fault tolerant control Two types of fault tolerant controllers can be designed. The active fault tolerant control, where the fault is accommodated when the fault is detected, and passive fault tolerant control, where the controllers are designed to stabilise the closed loop system in case of a fault. In this work only the active fault tolerant control is considered. Fault tolerant control methods normally consist of two parts which are:

- Fault detection and isolation (FDI): detects, isolates and estimates the fault.
- Fault accommodation: based on the detection, isolation and estimation of the fault, is the fault accommodated.

More information on fault accommodation and fault tolerant control is given in [Blanke et al., 2003].
Some examples of descriptions on fault detection and isolation are [Patton et al., 1989], [Gertler, 1998], [Frank, 1990] and [Chen and Patton, 1991]. The purpose of fault diagnosis is to diagnose the fault as early as possible before it develops and leads to severe consequences. In order to avoid these, this fault diagnosis consists of three parts: fault detection, fault isolation and fault estimation. In the normal usage the fault detection and fault isolation are essential in the fault diagnosis, these are often denoted FDI, on the other hand the fault estimation is not strictly necessary. However, in this application the fault detection and fault estimation are essential, and the fault isolation is not important, since the fault detection is used to detect a given kind of fault.
This leads to a fault detection and isolation (FDI) scheme consisting of the following three steps:

- Fault detection: is to detect the occurrence of a fault, typically indicated with a bit indicating the occurrence of a fault or not.
- Fault isolation: is to determine the origin of the fault.
- Fault estimation: is estimation of size and type of the fault, and of the fault itself.

Feature based control and fault tolerant control The fault detection/locator performs the fault detection, by locating the fault in time. The fault isolation in this situation is different from the standard fault isolation. Normally fault isolation is used to isolate the fault to be in a specific place in the system. In this application only surface faults are considered. By the construction of the decoupled residuals, the fault is isolated to be a surface fault. The estimation of the fault is performed by fault classification and approximation of the fault, where a disturbance decoupled fault estimation is performed. This

means that the fault detection together with the feature extraction give the fault diagnosis. These parts feed the fault accommodation with the information needed for handling the surface faults.
In Chapter 7, the surface faults were simulated by the use of Karhunen-Loève approximations. It is seen that just a few Karhunen-Loève basis vectors can be used to approximate the surface faults very well. It is an obvious idea to use the same basis vectors to approximate the surface fault, subtract that approximations from the measured detector signals the next time the fault is encountered. The algorithm will be described in the following section.

9.2 Fault accommodation by removal of the surface fault

It has previously been stated that a fault does not vary much from one encounter to the next encounter. This means that an approximation of the fault at encounter M can be subtracted from the fault at encounter $M+1$. This will remove the fault from the measured signals as well as it could if the approximation was used at encounter M. The signals used in the following are defined as follows: $\mathbf{e}[n]$ is a vector of focus and radial distances, $\check{\mathbf{e}}[n]$ is a vector of faulty sensor components due to the surface fault, $\tilde{\check{\mathbf{e}}}[n]$ is a vector of the estimates of the faulty sensor components due to the surface fault, $\grave{\check{\mathbf{e}}}[n]$ is a vector of the corrected sensor signals. $\mathbf{e}_{\mathrm{m}}[n]$ is a vector of the measured distance signals and $\mathbf{m}_{\mathrm{n}}[n]$ is a vector of the measurement noises. The relations among these signals are illustrated in Fig. 9.1.
Based on the previous chapters it is obvious to use the Karhunen-Loève basis, $\mathcal{K}$, to approximate the fault by using the q most approximating Karhunen-Loève basis vectors, $\mathbf{K}_{\check{e}}$. Consequently the remaining basis vectors, $\mathbf{K}_{\mathrm{n}}$, span the noises and disturbances in the distances signals.

$$\mathcal{K} = \begin{bmatrix} \mathbf{K}_{\mathrm{n}} & \mathbf{K}_{\check{e}} \end{bmatrix}, \tag{9.1}$$

where

$$\mathcal{K} \in \mathcal{R}^{(256\times 256)},\ \mathbf{K}_{\mathrm{n}} \in \mathcal{R}^{(256\times(256-q))},\ \mathbf{K}_{\check{e}} \in \mathcal{R}^{(256\times q)},\ a \ll 256. \tag{9.2}$$

Based on (9.1) the fault component in the measured distance signals can be separated from the measurement noises, which $\mathbf{K}_{\check{\mathbf{e}}}$ is constructed not to support. In addition an observation has been made by inspection of the measured data, that the disturbances give small projections on $\mathbf{K}_{\check{\mathbf{e}}}$. The reconstruction of measured signals without this surface will reduce the fault component of focus and radial distances dramatically.
An estimate of $\check{\mathbf{e}}[n]$ during a fault at encounter M can be computed by

$$\tilde{\check{\mathbf{e}}} = \begin{bmatrix} \mathbf{K}_{\check{e}_{\mathrm{f}}}\mathbf{K}_{\check{e}_{\mathrm{f}}}^{T} \cdot \left(e_{\mathrm{m,f}}^{L}[M] - \hat{e}_{\mathrm{f}}^{L}[M]\right) \\ \mathbf{K}_{\check{e}_{\mathrm{r}}}\mathbf{K}_{\check{e}_{\mathrm{r}}}^{T} \cdot \left(e_{\mathrm{m,r}}^{L}[M] - \hat{e}_{\mathrm{r}}^{L}[M]\right) \end{bmatrix}. \tag{9.3}$$

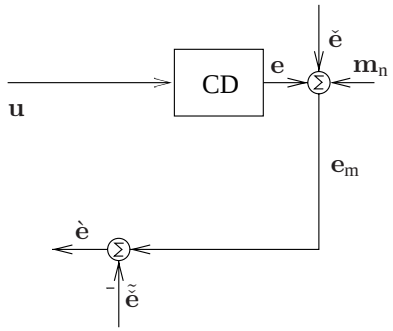


Figure 9.1: Illustration of the signals used in the feature based control scheme. CD is the CD-player. $\mathbf{u}$ is a vector of the control signals, $\mathbf{e}$ is a vector of focus and radial distances, $\check{\mathbf{e}}$ is a vector of faulty sensor components due to the surface fault, $\tilde{\check{\mathbf{e}}}$ is a vector of the estimates of the faulty sensor components due to the surface fault, $\grave{\mathbf{e}}$ is a vector of the clean distance signals. $\mathbf{e}_{\text{m}}$ is a vector of the measured distance signals and $\mathbf{m}_{\text{n}}$ is a vector of the measurement noises.

Here $\cdot^L[M]$ denotes the lifted signal, where the Mth fault encounter begins at sample no. n_M.
In the following a recurrent system is defined as a system where specific signal sequences of constant window lengths are recurring with none specific intervals. The length of these signals sequences are denoted l_w and the windows begins at samples $T_1, T_2 \cdots$.
A more detailed description of the lifting operator is given in [Wisniewski and Stoustrup, 2001] and [Khargonekar et al., 1985]. The lifting operator L is an isometric isomorphism which transforms a linear recurrent system to a time invariant representation defined as following

$$L : (y_0, y_1, \cdots)^T \mapsto \begin{bmatrix} (y_{T_1}, y_{T_1+1}, \cdots, y_{T_1+l_w-1}, \cdots) \\ (y_{T_2}, y_{T_2+1}, \ldots, y_{T_2+l_w-1}) \\ \vdots \end{bmatrix}^T , \tag{9.4}$$

where y is the signal which shall be lifted. In this case, the window length is equal m, since it is the length of the Karhunen-Loève basis vectors.
In order to use the approximation in (9.3), it is needed to estimate $\hat{\mathbf{e}}[n]$ by the use of a Kalman estimator described in Section 4.3 and [Odgaard et al., 2003a]. However, the use of this Kalman estimator can be avoided if $\hat{\mathbf{e}}[n]$ do not have large projections of $\mathbf{K}_{\check{\mathbf{e}}}$, since (9.3) can be approximated by (9.5).

$$\tilde{\check{\mathbf{e}}} = \begin{bmatrix} \mathbf{K}_{\check{e}_{\text{f}}} \mathbf{K}_{\check{e}_{\text{f}}}^T \cdot e_{\text{m,f}}^L[M] \\ \mathbf{K}_{\check{e}_{\text{r}}} \mathbf{K}_{\check{e}_{\text{r}}}^T \cdot e_{\text{m,r}}^L[M] \end{bmatrix} . \tag{9.5}$$

The Kalman estimator is only used to estimate $\hat{\mathbf{e}}[n]$, this means that a consequence of (9.5) is that the Kalman estimator is not needed in the closed loop feature based control scheme, but it is needed in order to compute the Karhunen-Loève basis of the surface faults. This means that if focus and radial distances are in the nominal operation range, the faults can be approximated by the use of the normalised focus and radial difference signals:

$$e_{\mathrm{m,f}}[n] \approx k_{\mathrm{f}} \cdot \frac{D_1[n] - D_2[n]}{D_1[n] + D_2[n]}, \tag{9.6}$$

$$e_{\mathrm{m,r}}[n] \approx k_{\mathrm{r}} \cdot \frac{S_1[n] - S_2[n]}{S_1[n] + S_2[n]}, \tag{9.7}$$

where k_{f} is the optical gain in the focus loop and k_{r} is the optical gain in the radial loop. Due to limitations in the PC computational powers only the approach with out the Kalman estimator is tested on the test setup. An extension of the proposed method could contain an adaptive scheme where the basis is trained based on faults on a given CD.

The stability and performance issues of the removal is dealt with in Section 9.3.

The approximation of the surface fault is now computed. The next problem is to determine when to begin the correction of focus and radial distances. This involves a synchronisation of the correction with the distance signals, where a correct synchronisation results in a removal of the fault from the measurements. An incorrect synchronisation might result in an increase of the controller reaction to the surface fault, and could actually make the problem with the surface faults more severe than if no correction was performed.

9.2.1 Synchronisation of the fault removal

In order to synchronise the correction of the measured signals two methods were used: detection of beginning and end of the surface fault, and prediction of the next fault based on previous encounters of the faults. The first method uses detection at a given fault to correct this given fault. This method has a good synchronisation, but unfortunately has the drawback that it locates the beginning and end of the faults respectively too late and too early. Using the previous location of the fault in time to predict the next fault, makes it possible to begin and end the correction at a more correct time than if detection based on the given fault is used. This prediction is based on some time localisation scheme e.g. the ones developed in Chapter 6 and [Odgaard and Wickerhauser, 2004]. However, it is not possible in practical settings to predict the placement of a fault. The reason for this is the implementation of the controller of the disc motor, which should guarantee a constant linear speed of the OPU relative to the track. However, this controller is implemented in a way that does not result in a constant linear speed but only in a linear speed in intervals.

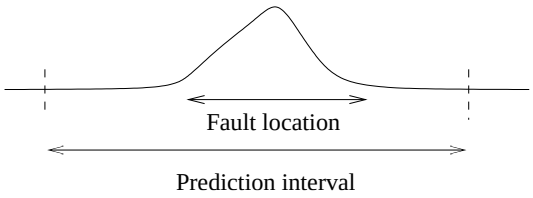


Figure 9.2: Illustration on the fault location and the interval in which the fault is located.

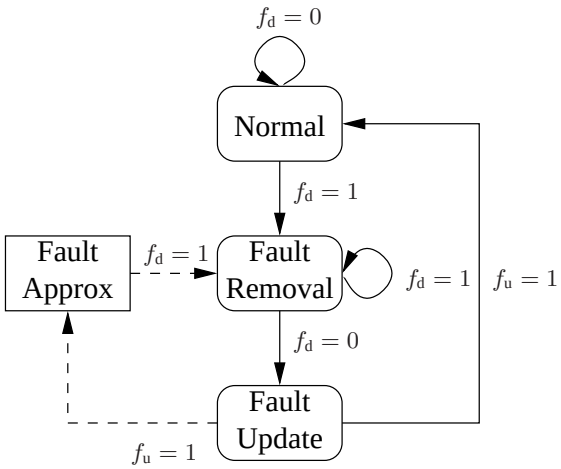


Figure 9.3: Illustration of the feature based control scheme illustrated as a state machine. The dashed lines do not illustrate state changes but data transfers. f_d denotes the detection signal, if a fault is detected at sample n it has 1 as its value, if a fault is not present it takes the value 0. f_u denotes an indication of the fault has been updated. The Fault Removal represents step 2 in the feature based control scheme and Fault Update represents step 3-5 in this scheme.

In practice a combination of the two fault localisation methods are used. The prediction, based on the time localisation methods presented in [Odgaard and Wickerhauser, 2004], is used to give an interval in which the fault is located and in this interval a lower threshold, than if the prediction was not used, can be used, see Fig. 9.2.

9.2.2 The algorithm of the feature based control strategy

The fault correction algorithm can now be stated:

1. Detect the fault and locate its position in time, when the fault is detected at sample n, $f_d[n] = 1$.

2. If $f_{\mathrm{d}}[n] = 1$:

$$a = \begin{cases} 0 & \text{if } f_{\mathrm{d}}[n-1] = 0, \\ a+1 & \text{if } f_{\mathrm{d}}[n-1] = 1. \end{cases},$$

$$\grave{\mathbf{e}}[n] = \mathbf{e}_{\mathrm{m}}[n] - \begin{bmatrix} \tilde{\check{\mathbf{e}}}_{\mathrm{f}}[\iota] \\ \tilde{\check{\mathbf{e}}}_{\mathrm{r}}[\iota] \end{bmatrix},$$

   where

$$\iota = ((256 - l_{\mathrm{f}}) \text{ div } (2)) + a,$$

   where a is a counter counting the number of samples the given fault is present, and ι is a counter used to locate the given sample relative to the fault correction block.

3. When the fault has been passed, classify the fault, time locate the fault, and compute the fault length l_{f}.

4. Compute the focus correction block coefficients by: $\mathbf{k}_{\mathrm{f}} = \mathbf{K}_{\check{e}_{\mathrm{f}}} \cdot e_{\mathrm{m,f}}[\upsilon]$ and the radial correction coefficients by : $\mathbf{k}_{\mathrm{r}} = \mathbf{K}_{\check{e}_{\mathrm{r}}} \cdot e_{\mathrm{m,r}}[\upsilon]$, where υ is the interval of 256 samples in which the fault is present.

5. Compute the focus fault removal correction block by: $\tilde{\check{\mathbf{e}}}_{\mathrm{f}} = \mathbf{K}_{\check{e}_{\mathrm{f}}} \cdot \mathbf{k}_{\mathrm{f}}$, and the radial fault removal correction by: $\tilde{\check{\mathbf{e}}}_{\mathrm{r}} = \mathbf{K}_{\check{e}_{\mathrm{r}}} \cdot \mathbf{k}_{\mathrm{r}}$.

In the following, $\mathcal{P}(\cdot)$ denotes an operator that maps measured signals into "repaired" ones by applying the correction algorithm. Due to the design of $\mathbf{K}_{\check{\mathbf{e}}}$ it does in principle not make any difference if $\mathbf{e}_{\mathrm{m}}[n]$ or $\mathbf{e}_{\mathrm{m}}[n] - \hat{\mathbf{e}}[n]$ is used to estimate the surface fault, since $\mathbf{K}_{\check{\mathbf{e}}}$ is designed to support $\check{\mathbf{e}}[n]$ and assumed not support $\mathbf{e}_{\mathrm{m}}[n]$. This means that

$$\mathcal{P}\left(\mathbf{e}_{\mathrm{m}}[n]\right) \approx \mathcal{P}\left(\check{\mathbf{e}}\right)[n] \Rightarrow \tag{9.8}$$

$$\mathcal{P}\left(\hat{\mathbf{e}}\right)[n] \approx 0. \tag{9.9}$$

Which in turn implies that it is not necessary to estimate $\hat{\mathbf{e}}[n]$. The normed focus and radial differences, (9.6, 9.7), can be used instead and thereby saves computing power. Only the last one is implemented due to limitations in the computer power in the experimental setup.

9.2.3 Practical implementation of the algorithm

The implementation uses only the four most approximating basis vectors, of a given fault class for each of two distance signals. The four basis vectors are chosen in order to limit the number of computations in the algorithm, and since experiments have shown that these four basis vectors approximate the faults very well see above. In order to avoid

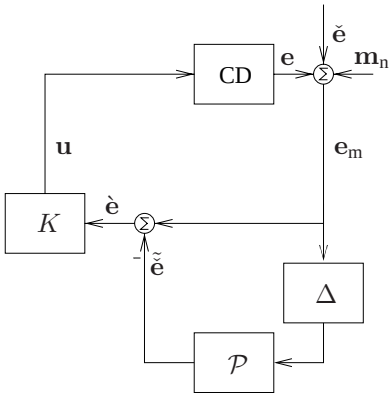


Figure 9.4: Illustration of the closed loop with the feature based correction $\mathcal{P}$, K is the controller, and CD is the CD-player. Δ is the one revolution delay. $\mathbf{u}$ is a vector of the control signals, $\mathbf{e}$ is a vector of focus and radial distances, $\check{\mathbf{e}}$ is a vector of faulty sensor components due to the surface fault, $\tilde{\check{\mathbf{e}}}$ is a vector of the estimates of the faulty sensor components due to the surface fault, $\grave{\mathbf{e}}[n]$ is a vector of the corrected sensor signals. $\mathbf{e}_{\text{m}}$ is a vector of the measured distance signals and $\mathbf{m}_{\text{n}}$ is a vector of the measurement noises.

book keeping algorithms it is assumed that the CD only has one scratch. This feature control scheme is scheme which accommodates the fault without changing the nominal controllers, in order to illustrate this fact the applied controllers are the frequently used PID-controllers see Chapter 8.2.3, Appendix D and [Stan, 1998].

9.3 Stability and performance of the feature based control scheme

The feature based control strategy is illustrated in Fig. 9.4. This figure illustrates how the influence from the surface fault is removed by the use of $\mathcal{P}(\cdot)$. This means that the controller, K, reacts on the sum of $\mathbf{e}[n]$ and the measurement noises, $\mathbf{m}_{\text{n}}$.
In the following some stability and performance issues of the algorithm will be discussed starting with dealing with the stability.

9.3.1 Stability

It is assumed that: K and CD are internally stable, and the nominal controller K stabilises the plant CD, if $\check{e}[n]$ is zero or near zero. However, if $\check{\mathbf{e}}[n]$ increases it might force the CD-player outside its linear region and could cause an unstable closed loop.

On the other hand if $\mathbf{e}[n] \approx \grave{\mathbf{e}}[n]$ the effect from the surface fault has been removed from the closed loop. This means that the control signal would be same as in the fault free case, meaning that the system is stable since it is nominally stable. $\mathcal{P}$ reconstruct the recurrent part of the measurement signals meaning that

$$\mathbf{e}_{\mathrm{m}} - \mathcal{P}\left(\mathbf{e}_{\mathrm{m}}\right) \approx= \mathbf{e} + \mathbf{m}_{\mathrm{n}}. \tag{9.10}$$

This can be achieved if

$$\check{\mathbf{e}} \approx \mathcal{P}\left(\mathbf{e}_{\mathrm{m}}\right) \Rightarrow \tag{9.11}$$

$$\mathcal{P}\left(\mathbf{m}_{\mathrm{n}}\right) \approx 0 \ \wedge \ \mathcal{P}\left(\mathbf{e}\right) \approx 0 \wedge \tag{9.12}$$

$$\check{\mathbf{e}} = \mathcal{P}\left(\check{\mathbf{e}}\right). \tag{9.13}$$

This is fulfilled if the approximating bases in $\mathcal{P}\left(\cdot\right)$ does approximate $\check{\mathbf{e}}[n]$ well and not $\mathbf{e}[n]$, $\mathbf{m}_{\mathrm{n}}[n]$. $\mathcal{P}\left(\cdot\right)$ is designed to be the best approximating basis of $\check{\mathbf{e}}[n]$. The Karhunen-Loève basis is not designed to support $\mathbf{e}[n]$, but no guarantees are given that it does not support $\mathbf{e}[n]$. Instead $\mathcal{P}\left(\cdot\right)$ might amplify the system dynamic in a degree that causes the entire system to be non-stable. From this it is clear that $\mathcal{P}\left(\cdot\right)$'s amplification of the system dynamic shall be small, such that the energy in a given system response is decreased through $\mathcal{P}\left(\cdot\right)$ from revolution to revolution. By inspecting Fig. 9.4 it can be seen that the way the feature based control scheme influences the control loop, such that influence can be analysed by using the complementary sensitivity of the servo system. The influence from the feature based fault handling on the nominal servo system can be inspected if the T denotes the complementary sensitivity of the nominal servo system, and Δ is the one revolution delay, see Fig. 9.5.
In order to combine these part systems, the complementary sensitivity of the nominal servo system and $\mathcal{P}$ are lifted, meaning that both part systems are represented by a discrete time series of a given length.
The lifted $\mathcal{P}$ can be computed by

$$\mathcal{P}^{\mathrm{L}} = \mathbf{K}_{\check{e}} \cdot \mathbf{K}_{\check{e}}^{T}, \tag{9.14}$$

and the lifted representation of the complementary sensitivity is

$$T^{\mathrm{L}} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & & \vdots \\ \vdots & & \ddots & \\ h_{255} & \cdots & & h_0 \end{bmatrix}, \tag{9.15}$$

where $\mathbf{h} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{255} \end{bmatrix}$ is time series of 256 samples of the impulse response of T.

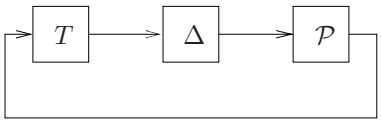


Figure 9.5: Closed loop of the feature based control system. T is the complementary sensitivity of the nominal system, Δ is the one revolution delay, and $\mathcal{P}$ is the feature based fault handling.

By lifting the system illustrated in Fig. 9.5 one gets a set of discrete differential equations of the form, if :

$$\xi[\upsilon+1] = \mathbf{A}\xi[\upsilon] + \mathbf{K}\mathbf{u}[\upsilon], \tag{9.16}$$

where $\mathbf{A} = T^{\mathrm{L}}\mathcal{P}^{\mathrm{L}}$. These definitions make it possible to state Theorem 9.1, which says that the linear system is stable.

Theorem 9.1 *The feature based control system defined by Fig. 9.4 is stable if and only if: max* $\left(|eig\left(T^{L}\mathcal{P}^{L}\right)|\right) < 1$,
where $\mathcal{P}^{L}$ *is defined in (9.14) and* T^{L} *is defined in (9.15).*

Proof:
Necessary and sufficient conditions:
The stability of the closed loop system shown in Fig. 9.4 is equivalent to stability of the system in (9.16), which is a standard LTI discrete time system, from which the result follows, the system is stable if and only if max $\left(|\text{eig}\left(T^{\mathrm{L}}\mathcal{P}^{\mathrm{L}}\right)|\right) < 1$. □
It is now possible to test if the feature based control scheme is stable regarding both focus and radial loops given the respective linear models. It is done by using models of the focus and radial loops, the nominal controllers and the computed $\mathbf{K}_{\check{e}_{\mathrm{f}}}$ and $\mathbf{K}_{\check{e}_{\mathrm{r}}}$. The computed value in the focus case is 0.6894 and in the radial case 0.0499. It is done with the conclusion that the linear stability criteria are fulfilled for both servo loops. I.e. as long the focus and radial servos are in the linear region these controllers are stable.

9.3.2 Performance of feature based control scheme

In order to inspect the performance of the algorithm it is needed to take the close loop into account. It can be done by determining the influence from the close loop on the approximation of the surface fault. In this regard influence will change from encounter to encounter due to the dynamics of the close loop, and since the system is linear and stable it will convergence. By inspecting Fig. 9.4, it can be seen that the approximation at encounter 1, $\mu = 1$.

$$\tilde{\check{\mathbf{e}}}_1^L = \mathcal{P}^{\mathrm{L}}\left(\mathbf{I} + \mathbf{S}^{\mathrm{L}}\right)\check{\mathbf{e}}_0^L, \tag{9.17}$$

where $\mathbf{S}^{\mathrm{L}}$ is the lifted sensitivity of the servo. At encounter $\mu = 2$ the approximation is

$$\tilde{\check{\mathbf{e}}}_2^L = \mathcal{P}^{\mathrm{L}}\left(\mathbf{I} + \mathbf{S}^{\mathrm{L}} - \mathbf{T}^{\mathrm{L}}\mathcal{P}^{\mathrm{L}}\mathbf{S}^{\mathrm{L}}\right)\check{\mathbf{e}}_1^L, \tag{9.18}$$

and at encounter $\mu = 3$

$$\tilde{\check{\mathbf{e}}}_3^L = \mathcal{P}^{\mathrm{L}} \left(\mathbf{I} + \mathbf{S}^{\mathrm{L}} - \mathbf{T}^{\mathrm{L}} \mathcal{P}^{\mathrm{L}} \mathbf{S}^{\mathrm{L}} \right) \check{\mathbf{e}}_2^L + \mathbf{T}^{\mathrm{L}} \mathcal{P}^{\mathrm{L}} \mathbf{T}^{\mathrm{L}} \mathcal{P}^{\mathrm{L}} \mathbf{S}^{\mathrm{L}} \check{\mathbf{e}}_3^L. \tag{9.19}$$

The influence from the closed loop can be determined by computing the energy of $\tilde{\check{\mathbf{e}}}_{1,2,3}$ over the energy of $\mathcal{P}^{\mathrm{L}} \check{\mathbf{e}}[n_\mu],\ \mu \in \{1,2,3\}$. I.e

$$\frac{\left\| \check{\mathbf{e}} - \tilde{\check{\mathbf{e}}} \right\|}{\left\| \check{\mathbf{e}} \right\|}. \tag{9.20}$$

These energy ratios are compute for the signals in the data set, and mean of all these ratios are following computed for focus and radial servos. The ratio for focus is 0.1134 and for radial 0.0033 for all the encounters meaning that, given this system and controller and the approximating basis and surface faults, the algorithm converge the first time it is applied. In addition it can be seen that the influence on the approximations from the controller and CD-player is small, meaning that the performance of the algorithm is almost only depending on the quality of the approximations of the surface faults, which has previous been concluded to be quit well.

9.4 Simulations of the Feature based controls scheme

In this section the feature based control scheme is simulated using the simulation model presented in Chapter 8. This simulation model simulates a CD with only one scratch of a given fault class. The standard PID-controllers are used in both focus and radial servo loops, see Chapter 8. Two simulations are presented, the first one uses $\mathcal{P}\left(\mathbf{e}_{\mathrm{m}}[n] - \hat{\mathbf{e}}[n]\right)$ to compute the estimate of the surface fault see Fig. 9.6, and in the second simulation $\mathcal{P}\left(\mathbf{e}_{\mathrm{m}}[n]\right)$ is used to compute the approximation, see Fig. 9.7. By comparing the figures it can be seen that in this simulation both methods correct the measurements in the way that the controller outputs do not have a component which is a response to the surface fault. The only visual difference is the number of revolutions it takes the two algorithms to remove the surface fault from the measurements.

Computing the $\mathcal{DP}$, see Definition 8.1, of the two simulations gives a clear result that the first method has the best performance handling the surface fault. However, seen from Figs. 9.6 and 9.7 it seems that the second method ends with the same fault handling performance. In order to compare the methods "steady-state" fault handling performance, the $\mathcal{DP}$ is computed for the last scratch in the simulation for both the methods. These results can be seen in Table 9.1. From this table it can be seen that the first method is better than the second method on $e_{\mathrm{f}}[n]$, and the second method is only marginally better on $e_{\mathrm{f}}[n]$. This is properly due to the fact that measurement noises are relatively larger compared with the faults in the radial signals than in the focus signals.

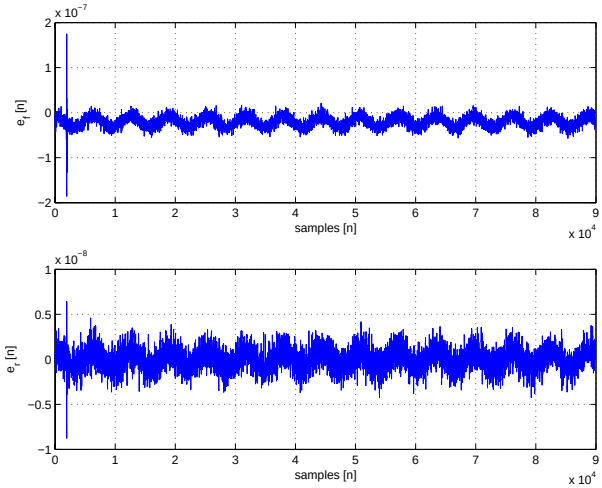


Figure 9.6: Simulations of the fault accommodation based on $\mathbf{e}_{\mathrm{m}}[n] - \hat{\mathbf{e}}[n]$ handling a scratch. The upper plot is the focus distance, and the lower plot is the radial distance. The first encounter of the fault is used to extract features from the scratch meaning that no correction is performed.

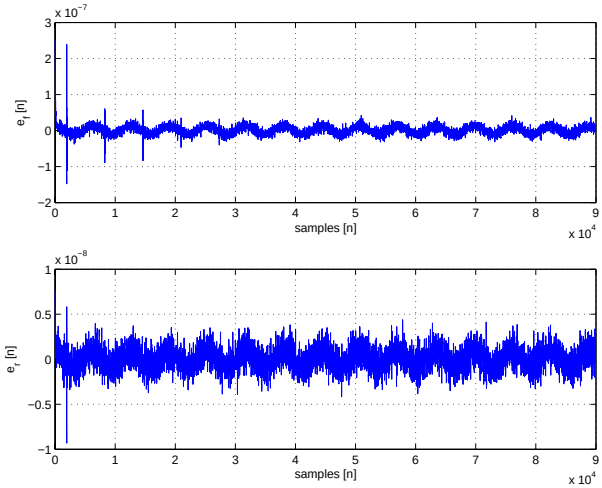


Figure 9.7: Simulations of the fault accommodation based on $\mathbf{e}_{\mathrm{m}}[n]$ handling a scratch. The upper plot is the focus distance, and the lower plot is the radial distance. The first encounter of the fault is used to extract features from the scratch meaning that no correction is done.

	The first method	The second method
$\mathcal{DP}$	$\begin{bmatrix}0.6416\\0.4318\end{bmatrix}$	$\begin{bmatrix}1.9023\\0.4468\end{bmatrix}$

Table 9.1: Table of the fault handling performance $\mathcal{DP}$, for both the feature based control methods. It can be seen that the first method is better than the second method on $e_{\mathrm{f}}[n]$, but the second method is marginally better on $e_{\mathrm{f}}[n]$. This is properly due to the fact that measurement noises are relatively larger compared with the faults in the radial signals than in the focus signals.

9.5 Experimental results

The next step is to verify the algorithm on the experimental test setup, see Chapter 3.1. In this experimental work only CDs with one scratch has been used, in order to avoid book keeping of the surface faults. However, this scratch has not been included in the training set of the algorithm. The scratch has also been classified to be contained in Class 1 - small scratches. Due to limitations in the computer power in the test setup, the algorithm was not used on both focus and radial loops at the same time. For the same reason only the second method was tested with success. It has not been possible to implement the first method due to real-time problems. However, the other parts of the method suggested in this thesis have previously been validated by experimental work presented in this thesis. A scheme combining all these scheme parts is assumed to give better experimental results as simulations indicates.

In the first experiment the fault was detected by the use of prediction of the time localisation based on the location of the previous faults. A clear advantage of this approach is that the fault can be removed from the beginning of the surface fault, which should result in an almost entire removal of the scratch from the measured $\mathbf{e}_{\mathrm{m}}[n]$. This approach is highly sensible to non-deterministic variations in the length of the revolutions in samples, (variations mean additional variations to the increase in the duration of one revolution with one sample for each third revolution due to the spiral shaped track on the disc). Analyses of the experimental data show that the variations in the duration of the revolutions vary with up to 5 samples from one revolution to the next one. Such a large variation will result in problems for the fault correction algorithm. It might cause the use of the algorithm to result in a poorly performance than if the correction was not performed. The same analyses of the experimental data also show that the length of surface faults do not change from revolution to revolution, this fact is of large importance later on.

A couple of experiments were performed using the prediction of the fault location approach. These experiments resulted as presumed in a good fault correction when location prediction is correct, and on the other hand resulted in a very poorly fault correction when this prediction did fail, see Fig. 9.8.

In Fig. 9.8 the potentials of the algorithm are shown at the 4th and the 5th encounter

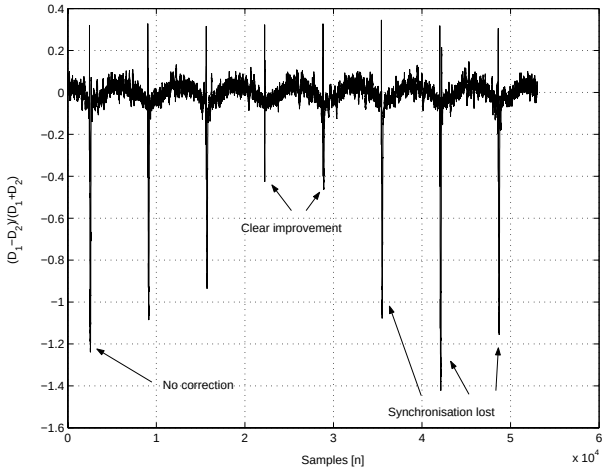


Figure 9.8: A plot of the measured normalised focus difference, while the feature based correction algorithm is applied with the use of prediction of the fault location. The fault accommodation of 4th and 5th encounters of fault results in a clear improvement. Unfortunately the fault correction in the following encounters is out of synchronisation. The correction results in the best case in no improvements.

of the surface faults, where the oscillation due to the fault in $e_{\mathrm{f}}[n]$ is clearly decreased compared with the first encounter which is a correction free encounter. Unfortunately in the following encounters the location prediction fails, which result in a poorly correction of the fault, and the correction results in no improvements at all.

The prediction of the time location of the fault is used to give a region in which the fault is present. A threshold algorithm is used inside this region to locate the fault. This threshold can be chosen lower than if the prediction was not used to given a region in which the fault is located. The fault correction is following applied when the fault is located and detected. It is assumed that the length of the fault is the same as at the last encounter, the length of the fault is denoted, f_{l}. The length of the correction block is 256, since $f_{\mathrm{l}} < 256$ for the entire dataset. As long as the fault is located, the correction can be written as

$$a[n] = \begin{cases} 0 & \text{if } f_{\mathrm{d}}[n-1] = 0, \\ a[n]+1 & \text{if } f_{\mathrm{d}}[n-1] = 1. \end{cases}, \tag{9.21}$$

$$\grave{\mathbf{e}}[n] = \mathbf{e}_{\mathrm{m}}[n] - \begin{bmatrix} \breve{\mathbf{e}}_{\mathrm{f}}[\iota] \\ \breve{\mathbf{e}}_{\mathrm{r}}[\iota] \end{bmatrix}, \tag{9.22}$$

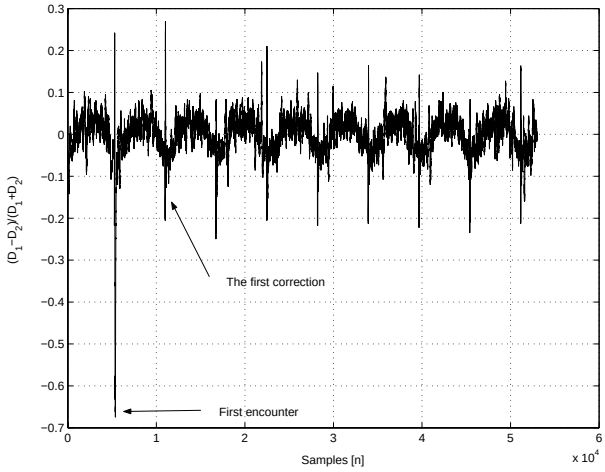


Figure 9.9: A plot of the measured normalised focus difference while the correction algorithm is applied with the use of fault detection. It is clear to see that the correction results in a limitation of the influence from the fault on the measured normalised focus difference.

where

$$\iota = ((256 - l_{\mathrm{f}}) \ \mathrm{div} \ (2)) + a. \tag{9.23}$$

In Fig. 9.9 an example of this method is illustrated. The first encounter of the scratch is not corrected, but instead used to train the algorithm. In the encounters following the first one a clear improvement is achieved. It can also be seen more clear from the zoom on the 1st and 5th encounter of the fault, see Fig. 9.10.
Following the algorithm was tested on the radial servo loop, which is known as being more sensitive to faults than the focus servo. This test was performed with the CD-player in test mode, see Chapter 3.1. An example of the algorithm handling a scratch can be seen in Fig. 9.11. The nominal controller actually looses the radial tracking in the example in Fig. 9.11 which can be seen by the heavy oscillations. However, when the correction algorithm is applied the radial tracking is not lost. A zoom on the 2nd and 4th encounter of the fault is shown in Fig. 9.12.
These experimental results show that a limited version of feature based fault correction algorithm, which is proposed in this chapter, gives a clear improvement of the performance handling the surface faults, at least in the case of the tested scratches. The entire algorithm might even give better results, as simulations indicates. In Section 3.2 it was stated: “The main problem to solve in this Ph.D project is to handle surface faults that are impossible to handle by normally used methods.” The feature based control scheme has not been tested on such surface faults, due to some practical problems. However, it

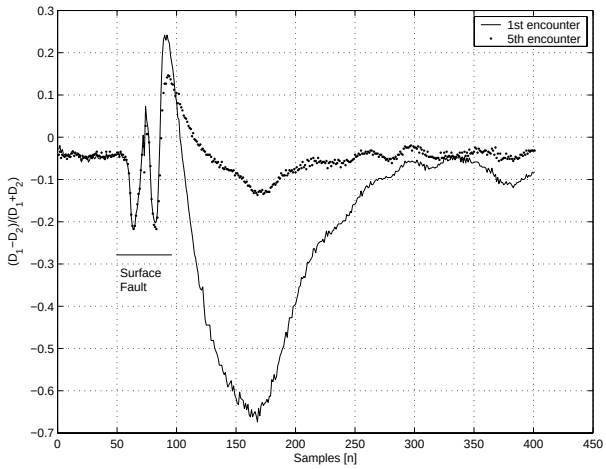


Figure 9.10: A zoom on the 1st and 5th encounter of the fault shown in Fig. 9.9.

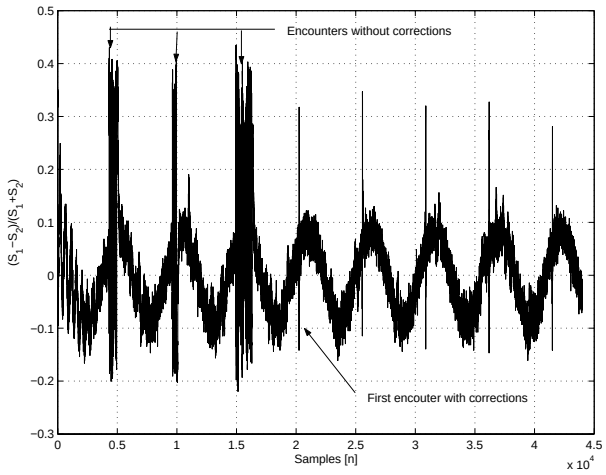


Figure 9.11: A plot of the measured normalised radial difference while the fault correction algorithm is applied with the use of thresholding in a predicted interval. It is clear to see that the fault correction algorithms result in limitation of the influence from the fault on the approximated $e_{\mathrm{r}}[n]$, when the correction algorithm is applied at the 4th encounter of the fault. It can be seen that if no correction is applied the fault forces the OPU into heavy oscillations, which are avoided when the correction algorithm is used.

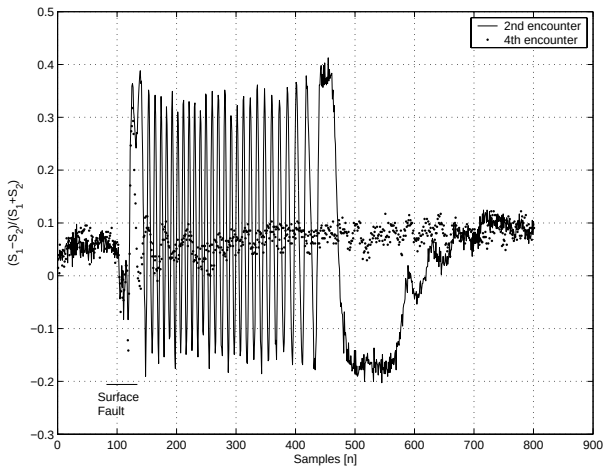


Figure 9.12: A zoom on the 2nd and 4th encounter of the fault shown in Fig. 9.11.

is experienced that if the controller reaction on surface faults in general is minimised, it is possible to play CDs with large surface faults. Based on this argument it can be concluded that the proposed feature based control scheme can improve the handle of CDs with surface faults, which cannot be played by the nominal controllers.

9.6 Summary

In this chapter the feature based control scheme has been derived and verified. In Section 9.1 this scheme is described in general terms and it is argued that the feature based control algorithm is significantly related to the fault tolerant control method. This leads to the fault accommodation part of the feature based control algorithm which is described in Section 9.2. The algorithm is based on removal of the surface fault components in the distance signals. In order to do so the surface faults are approximated by computing the projection of the normalised measured focus and radial differences feedback on the Karhunen-Loève bases approximating the surface faults. The estimated fault free distance signals are computed by subtracting the estimated surface fault from the measured distance signals. Some stability and performance issues of the feature based control algorithm are discussed in Section 9.3, and it is proven that the scheme is stable given the linear model of the servo system. In Section 9.4 verification of the algorithm is performed by the use of the simulation model derived in Chapter 8. These simulations show that the feature based control scheme improves the accommodation of the surface fault. The same conclusion is achieved by inspecting the experimental results presented in Section 9.5, where a limited version of the feature based control scheme is used.

Chapter 10

Conclusion

This thesis deals with different aspects on how to improve the playability of the CD-players for CDs with surface faults, such as scratches and fingerprints. The suggested method called feature based control, is in its essential an active fault tolerant control method combined with time-frequency based signal processing. In the following the contributions of the thesis are summarised and reviewed. This summary is followed by some suggestions for future work, which is based on this project.

10.1 Conclusions

In this thesis fault tolerant control is used for a class of systems, where repetitive sensor faults occur. This class of systems contains applications in the field of optical disc players playing discs with surface faults. A scheme is developed which is adapted to a specific CD-player. However, it is expected that this scheme can be used in general for for system in the considered class of systems. This novel method is based on a number of methods and results which are achieved in the areas of: modelling of the optical system, design of residuals, fault detection, time-frequency based feature extraction, simulations and feature based control.

- A model of the optical detector system is derived. This model includes the optical cross-couplings between focus and radial detectors. Outside the normal operation range of the focus and radial loops this optical mapping is strongly non-linear. The derived model is validated based on measurements. The validation shows that the model is well suited for control purposes. However, the derived model has some numerical problems which is handled by approximating the functions in the model by the use of cubic splines.

- In most industrial players either the focus or the radial sum signal is used as the residual for detection of the surface faults. Unfortunately the focus sum signal

depends the radial distance and the radial sum signal depends on the focus distance. Instead a couple of decoupled residuals are derived based on a model of the surface faults. These residuals are compared with the normally used residuals on some real scratches and show a clear improvement of the localisation of these scratches.

- The previously mentioned residuals can only be computed if the inverse map of the combined optical and fault model is used. This mapping is only locally defined computational method is suggested by the use of Newton-Raphson's method and a Kalman estimator. In addition the Kalman estimator gives an estimate of focus and radial distances during a surface fault.

- The decoupled residuals are the starting point for the fault detection and localisation. Three methods are designed to locate and detect the surface faults based on these residuals. Two time-frequency based methods are designed, and as well as an extension of the normally used threshold. The two time-frequency based methods are Fang's algorithm, and a wavelet packet filter. These two methods has the drawback that one filter cannot be used to detect all surface faults. Instead a set of filters are needed in order to locate all surface faults by the use of these wavelet packet filters. However, the extended threshold method gives a clear improvement of the fault detection and localisation in all tested cases.

- Time-frequency based methods are used to extract features from the detected surface faults. The extracted features are classification of the surface faults, (this classification separates the faults in smaller classes thereby more adapted fault handling can be achieved), and approximating coefficients in the Karhunen-Loève basis. These features are used for two purposes, simulation of the surface fault, and to accommodate them by removal of the surface faults component from the measurements. The classification and approximation are verified on test data which are not used in the data set to train the classification and approximation.

- In order to test how well a controller can handle a surface fault, a simulator of surface faults is designed by using the approximating Karhunen-Loève basis. These synthesised surface fault are simulation of specific surface faults, but contained in the given class of surface faults. The simulation model is completed by the use of standard models of focus and radial servos combined with the new optical model developed in this thesis.

- The feature based control method is designed to handle the surface faults. By subtracting the estimation of surface fault from focus and radial distance signals. These cleaned distance signals are subsequently fed to the nominal controllers. The nominal controller can be used without reacting to the surface faults. The method is in this design limited to handle only one scratch on the played disc. This method is validated through simulations and experimental tests, where a simplified version was tested with good results. In these simulations and tests the often

used PID-controllers were used in order to illustrate the fact that the feature based control scheme accommodates the surface faults without using special controllers.

All these novel contributions lead to a method which can improve the performance of the focus and radial servos in a CD-player, playing a CD with a surface faults, such as scratches and fingerprints. The experimental work has been focused on the scratches and the method is as a consequence adapted to the scratches. The method as one has not been tested by the experimental tests. However, the parts of the methods have been tested by experimental work with good results. The feature based control method used on a CD-player has both in simulations and experimental tests (a simple version) shown improvements in the handling of the tested scratches.

10.2 Future work

Several ideas and problems have appeared but not covered during this Ph.D project. These problems and ideas will be described in the following.

- This feature based method is designed to a specific CD-player. It would be very interesting to extend the method to other optical disc players. The obviously choice of the first player to extend it to, is the DVD-player, but also the new high density standards are of high interest. In addition it would also be interesting to adapt the method to high speed drives.
- The fault classification and estimation in the proposed method are based on a limited training set of measured surface faults. This means that surface faults exist which are poorly represented by the fault classes and approximating bases. This will result in a poor handling of the surface faults. One way to improve this is to expand the training set.
- Only a small number of bases were tested in the fault classification method, other bases might be even more discriminating than the one used. Another possibility is to combine different libraries of bases, meaning that a set of base vectors from different bases are chosen.
- The large number of surface faults can be handled in an other way. The first encounter of the fault is used to train an approximating Karhunen-Loève basis of the surface fault. This method has an even larger potential if the CD-player has the storage capacity to store the features of the surface faults on all the CDs it has played. It is probably necessary, due to limited storage capacity, to store only the most problematic surface faults on a number of CDs.
- It might also be useful to adapt the fault detection filters to the classified faults. This would make the use of the time-frequency based filters possible, since these filters tend to be too narrow-banded to be used for all faults.

- It was not possible to implement the entire algorithm due to real-time problems in the test setup. As a consequence it would be a good idea to design a faster version of the method. One of the problems is the computation of the inverse map. Some studies have indicated that it is possible to determine the inverse of the model mapping locally, which can remove all the iterative computations.
- The method as it is now can only handle one surface fault at each CD. As a consequence it would be a good idea to implement a book keeping algorithm which makes it possible to handle more than one surface fault per CD.
- Another possible usage of this method is the reconstruction of data on severe damaged discs, where playback is not possible even though the suggested method is used. It might be possible to retrieve the data from the disc by changing the rotational speed of the disc in addition to the use of the suggested method.
- A final interesting topic to investigate further is how surface faults influence focus and radial distance measurements. This might be achieved by studying diffractive theory.

Bibliography

[Aangenent, 2002] Aangenent, W. (2002). The modeling and control of the sledge system in the Beosound 3000. Technical report, Department of Control Engineering, Institut of Electronic Systems, Aalborg University, Denmark.

[Andersen and Karlsson, 2000] Andersen, B. and Karlsson, R. (2000). Regulering af optisk pickup i cd-afspiller. Master's thesis, Department of Control Engineering, Institute of Electronic systems, Aalborg University, Denmark. In Danish.

[Andersen et al., 2001] Andersen, P., Pedersen, T., Stoustrup, J., and Vidal, E. (2001). Method for improved reading of digital data disc. International patent, no. WO 02/05271 A1.

[AOSRA, 2003] AOSRA (2003). Taiwan joins optical-disk IP fray. http://www.eetimes.com/sys/news/OEG20030324S0032.

[Basseville and Nikiforov, 1993] Basseville, M. and Nikiforov, I. (1993). *Detection of abrupt changes : theory and application*. Prentice Hall, first. edition.

[Blanke et al., 2003] Blanke, M., Kinnaert, M., Lunze, J., and Staroswiecki, M. (2003). *Diagnosis and Fault-tolerant Control*. Springer.

[Bouwhuis et al., 1985] Bouwhuis, W., Braat, J., Huijser, A., Pasman, J., van Rosmalen, G., and Schouhamer Immink, K. (1985). *Principles of Optical Disc Systems*. Adam Hilger Ltd.

[Braat et al., 2002] Braat, J., Dirksen, P., and Jansen, A. (2002). Diffractive readout of optical discs. In Török, P. and Kao, F.-J., editors, *Optical Imaging and Microscopy Techniques and Advanced Systems*. Springer, first edition.

[Chait et al., 1994] Chait, Y., Soo Park, M., and Steinbuch, M. (1994). Design and implementation of a QFT controller for a compact disc player. In *Proceedings of the 1994 American Control Conference*, Baltimore, Maryland, USA.

[Chen and Patton, 1991] Chen, J. and Patton, R. J. (1991). *Robust model-based fault diagnosis for dynamic systems*. Kluwer academic publishers, first edition.

[Coifman and Saito, 1994] Coifman, R. R. and Saito, N. (1994). Constructions of local orthonormal bases for classification and regression. *Comptes Rendus de l'Académie des Sciences de Paris*, 319(2):191–196.

[Coifman and Wickerhauser, 1992] Coifman, R. R. and Wickerhauser, M. V. (1992). Entropy-based algorithm for best basis selection. *IEEE Transaction on Information Theory*, 38(2):713–718.

[Dettori, 2001] Dettori, M. (2001). *LMI techniques for control - with application to a Compact Disc player mechanism*. PhD thesis, Technische Universiteit Delft, The Netherlands.

[Dettori et al., 1999] Dettori, M., Valk, P., and Scherer, C. (1999). Digital implementation of a mixed objectives mimo controller for a compact disc player using a multiprocessor system. In *Proceedings of the 1999 American Control Conference*, San Diego, CA, USA.

[Dokur et al., 1999] Dokur, Z., Ölmez, T., and Yazgan, E. (1999). Comparison of discrete wavelet and fourier transforms for ecg beat classification. *Electronics Letters*, 35(18):1502 – 1504.

[Dötch et al., 1995] Dötch, G., Smakman, T., Van den Hof, P., and Steinbuch, M. (1995). Adaptive repetitive control of a compact disc mechanism. In *Proceedings of the 34th IEEE Conference on Decision and Control*, New Orleans, LA ,USA.

[Draijer et al., 1992] Draijer, W., Steinbuch, M., and Bosgra, O. (1992). Adaptive control of the radial servo system of a compact disc player. *Automatica*, 28(3):455–462.

[Dzanovic and Lauritsen, 2003] Dzanovic, H. and Lauritsen, K. (2003). Modellering af slædesystemet i et cdm12 løbeværk. Technical report, Department of Control Engineering, Institute of Electronic Systems, Aalborg University, Denmark. In Danish.

[Dötsch, 1998] Dötsch, H. (1998). *Identification of Control Design with Application to a Compact Disc Mechanism*. PhD thesis, Technische Universiteit Delft, The Netherlands.

[ECMA, 2001] ECMA (2001). *Standard ecma-267: 120mm dvd read-only-disk*. Standard.

[Fang, 1994] Fang, X. (1994). Automatic phoneme segmentation of continuous speech signals. Preprint.

[Filardi, 2003] Filardi, G. (2003). *Robust control design strategies applied to a DVD-video player*. PhD thesis, Université Joseph Fourier, Grenoble, France.

[Filardi et al., 2003] Filardi, G. Sename, O., Besaçon-Voda, A., and Schroder, H. (2003). Robust $\mathcal{H}_\infty$ control of a dvd drive under parametric uncertainties. In *Proceedings of European Control Conference 2003*, Cambridge, UK.

[Flury, 1997] Flury, B. (1997). *A first course in multivariate statistics*. Springer Verlag, first edition.

[Frank, 1990] Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. a survey and some new results. *Automatica*, 26(3):459–474.

[Franklin et al., 1998] Franklin, G. F., Powell, J. D., and Workman, M. (1998). *Digital control of dynamic systems*. Addison Wesley, third edition.

[Fujiyama et al., 1998] Fujiyama, K., Tomizuka, M., and Katayama, R. (1998). Digital tracking controller design for cd player using disturbance observer. In *International Workshop on Advanced Motion Control*, Coimbra, Portugal.

[Gamba and Marazzi, 1997] Gamba, P. and Marazzi, A. (1997). A comparison between wavelet and threshold detection of rain patterns from radar data. In *Sixth International Conference on Image Processing and Its Applications*, Dublin, Ireland.

[Gertler, 1998] Gertler, J. J. (1998). *Fault Detection and Diagnosis in Engineering System*. Marcel Dekker, Inc.

[Goossens and Odgaard, 2003] Goossens, H. and Odgaard, P. (2003). Optical disc servo that is robust for defects. European patent application no. 03104587.5.

[Hearns and Grimble, 1999] Hearns, G. and Grimble, M. (1999). Limits of performance of an optical disk drive controller. In *Proceedings of the 1999 American Control Conference*, San Diego, CA, USA.

[Heertjes and Sperling, 2003] Heertjes, M. and Sperling, F. (2003). A nonlinear dynamic filter to improve disturbance rejection in optical storages drives. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, Maui, HI, USA.

[Heertjes and Steinbuch, 2004] Heertjes, M. and Steinbuch, M. (2004). Stability and performance of a variable gain controller with application to a dvd storage drive. *Automatica*, 40(4):591–602.

[Hess-Nielsen and Wickerhauser, 1996] Hess-Nielsen, N. and Wickerhauser, M. (1996). Wavelets and time-frequency analysis. *Proceedings of The IEEE*, 84(4):523 – 540.

[Hitachi, Ltd. et al., 2002] Hitachi, Ltd., LG Electronics Inc., Matsushita Electric Industrial Co., Ltd, Pioneer Corporation, Royal Philips Electronics, Samsung Electronics Co., Ltd., Sharp Corporation, Sony Corporation, and Thomson Multimedia (2002). Large capacity optical disc video recording format “blu-ray disc” established. Press Release: http://www.sony.net/SonyInfo/News/Press/200202/02-0219E/.

[Hnilička et al., 2002] Hnilička, B., Besançon-Voda, A., Filardi, G., and Schroeder, H. (2002). Modelling the focus error characteristic of a dvd player. In *Proceedings of the 2002 IEEE International Conference on Control Applications*, Glasgow, Scotland, UK.

[Honguh, 2002] Honguh, Y. (2002). Readout signal analysis of optical disk base don an approximated vector diffraction theory. In *International Symposium on Optical Memory and Optical Data Storage Topical Meeting 2002*, Waikoloa, HI, USA.

[Huang et al., 1999] Huang, Y., Mathur, P., and Messner, W. (1999). Robustness analysis on a high bandwidth disk drive servo system with an instrumented suspension. In *Proceedings of the 1999 American Control Conference*, San Diego, CA, USA.

[Jensen and la Cour-Harbo, 2001] Jensen, A. and la Cour-Harbo, A. (2001). *Ripples in Mathematics - The Discrete Wavelet Transform*. Springer, Berlin Heidelberg, Germany, first edition.

[Johnson and Wichern, 2002] Johnson, R. and Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall, 5th edition.

[Khargonekar et al., 1985] Khargonekar, P., Poolla, K., and Tannenbaum, A. (1985). Robust control of linear time-invariant plants using periodic compensation. *IEEE Transaction on Automatic Control*, 30(11):1088 – 1096.

[Kojima and Wakabayashi, 1997] Kojima, T. and Wakabayashi, K. (1997). Three-dimensional analysis of light-beam scattering from pit and emboss marks of optical disk models. In *1997. IEEE. Digest. Antennas and Propagation Society International Symposium*, Montreal, Que., Canada.

[Kraut et al., 2004] Kraut, S., Andersen, R., and Krolik, J. (2004). A generalized karhunen-loeve basis for efficient estimation of tropospheric refractivity using radar clutter. *IEEE Transaction on Signal Processing*, 52(1):48–60.

[la Cour-Harbo, 2002] la Cour-Harbo, A. (2002). *Robust and Low-Cost Active Sensors by means of Signal Processing Algorithms*. PhD thesis, Department of Control Engineering, Aalborg University. ISBN: 87-90664-13-2.

[Lee, 1998] Lee, C. H. (1998). *Robust repetitive control and application to a CD player*. PhD thesis, Trinity Hall Cambridge University.

[Li et al., 1995] Li, C., Zheng, C., and Tai, C. (1995). Detection of ecg characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1):21 – 28.

[Li and Tsao, 1999] Li, J. and Tsao, T.-C. (1999). Rejection of repeatable and non-repeatable disturbances for disk drive actuators. In *Proceedings of the 1999 American Control Conference*, San Diego, CA, USA.

[Lopez et al., 1994] Lopez, J. E., Tenney, R. R., and Deckert, J. C. (1994). Fault detection and identification using real-time wavelet feature extraction. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Philadelphia, PA, USA.

[Mallat, 1989] Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11(7):674 – 693.

[Mallat, 1999] Mallat, S. (1999). *A wavelet tour of signal processing*. Academic Press, second. edition.

[Mokhtari et al., 2002] Mokhtari, H., Karimi-Ghartemani, M., and Iravani, M. R. (2002). Experimental performance evaluation of a wavelet-based on-line voltage detection method for power quality applications. *IEEE Transaction on Power Delivery*, 17(1):161 – 172.

[Moon et al., 1996] Moon, J., Lee, M., and Jin Chung, M. (1996). Track-following control for optical disk drives using an iterative learning scheme. *IEEE Transactions on Consumer Electronics*, 42(2):192–198.

[Odgaard et al., 2003a] Odgaard, P., Stoustrup, J., Andersen, P., and Mikkelsen, H. (2003a). Detection of surface defects and servo signal restoration for a compact disc player. Submitted for journal publication.

[Odgaard et al., 2003b] Odgaard, P., Stoustrup, J., Andersen, P., and Mikkelsen, H. (2003b). Extracting focus and radial distances, fault features from cd player sensor signals by use of a kalman estimator. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, Maui, Hawaii, USA.

[Odgaard et al., 2003c] Odgaard, P., Stoustrup, J., Andersen, P., and Mikkelsen, H. (2003c). Modelling of the optical detector system in a compact disc player. In *Proceedings of the 2003 American Control Conference*, Denver, USA.

[Odgaard et al., 2004a] Odgaard, P., Stoustrup, J., Andersen, P., and Mikkelsen, H. (2004a). Fault detection for compact disc players based on redundant and non-linear sensors. In *Proceedings of American Control Conference 2004*, Boston, MA, USA.

[Odgaard et al., 2004b] Odgaard, P., Stoustrup, J., Andersen, P., and Wickerhauser, M. (2004b). Joint best basis wavelet packet based detection of surface faults on compact disc. Submitted for publication.

[Odgaard et al., 2003d] Odgaard, P., Stoustrup, J., Andersen, P., Wickerhauser, M., and Mikkelsen, H. (2003d). Matlab source for simulation of cd player with surface defects. www.control.aau.dk/∼odgaard/software/CDsim. The link includes matlab and data files for download, which are needed for using the simulation software.

[Odgaard et al., 2004c] Odgaard, P., Stoustrup, J., Andersen, P., Wickerhauser, M., and Mikkelsen, H. (2004c). A fault tolerant control scheme for cd players to handle surface defects. Submitted for journal publication.

[Odgaard et al., 2004d] Odgaard, P., Stoustrup, J., Andersen, P., Wickerhauser, M., and Mikkelsen, H. (2004d). A simulation model of focus and radial servos in compact disc players with disc surface defects. In *Proceedings of the 2004 IEEE CCA/ISIC/CACSD*, Taipei, Taiwan.

[Odgaard et al., 2004e] Odgaard, P., Stoustrup, J., Andersen, P., Wickerhauser, M., and Mikkelsen, H. (2004e). Stability of close loop controlled repetitive periodic systems. Submitted for publication.

[Odgaard and Wickerhauser, 2003] Odgaard, P. and Wickerhauser, M. (2003). Discrimination between different kind of surface defects on compact discs. To appear in Proceedings of IEEE IECON 2004.

[Odgaard and Wickerhauser, 2004] Odgaard, P. and Wickerhauser, M. (2004). Time localisation of surface defects on optical discs. In *Proceedings of the 2004 IEEE CCA/ISIC/CACSD*, Taipei, Taiwan.

[Ouyang and Wickerhauser, 1993] Ouyang, H. and Wickerhauser, M. V. (1993). WSQ – the FBI/Yale/Los Alamos [W]avelet-packet [S]calar [Q]uantization fingerprint compression algorithm, for Windows 3.1 or higher. http://www.math.wustl.edu/∼victor/software/wsq/index.html.

[Parameswariah and Cox, 2002] Parameswariah, C. and Cox, M. (2002). Frequency characteristics of wavelets. *IEEE Transactions on Power Delivery*, 17(3):800 – 804.

[Patton et al., 1989] Patton, R., Frank, P., and Clark, R. (1989). *Fault Diagnosis in Dynamic Systems - Theory and Application*. Prentice Hall International.

[Philips, 1994] Philips (1994). *Product specification: Digital servo processor DSIC2, TDA1301T*. Philips Semiconductors, Eindhoven, The Netherlands.

[Philips and Sony Corporation, 1991] Philips and Sony Corporation (1991). Compact disc digital audio. Red book, which is non public standard, but most of the relevant specifications can be seen in other literature e.g. [Stan, 1998].

[Pohlmann, 1992] Pohlmann, K. (1992). *The Compact Disc Handbook*. Oxford University Press, 2nd edition.

[Reed and Solomon, 1960] Reed, I. and Solomon, G. (1960). Polynomial codes over certain finite fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304.

[Saito, 1994] Saito, N. (1994). *Local Feature Extraction and Its Applications Using a Library of Bases*. PhD thesis, Yale University.

[Saito, 1998] Saito, N. (1998). The least statically-dependent basis and its applications. In *Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA.

[Schneiders, 2001] Schneiders, M. (2001). Wavelets in control engineering. Master's thesis, Eindhoven University of Technology. Dynamics and Control Technology, Faculty of Mechanical Engineering, Eindhoven University of Technology.

[Stan, 1998] Stan, S. G. (1998). *The CD-ROM drive*. Kluwer Academic Publishers.

[Steinbuch et al., 1992] Steinbuch, M., Schoostra, G., and Bosgra, O. (1992). Robust control of a compact disc player. In *Proceedings of the 31st IEEE Conference on Decision and Control*, Tucson, Arizona, USA.

[Steinbuch et al., 1994] Steinbuch, M., van Groos, P., Schootstra, G., and Bosgra, O. (1994). Multivariable control of compact disc player using DSPs. In *Proceedings of the American Control Conference*, Baltimore, Maryland, USA.

[Strang and Nguyen, 1997] Strang, G. and Nguyen, T. (1997). *Wavelets and Filter Banks*. Wellesley- Cambridge Press, Wellesley, MA, revised edition.

[Sureshbabu and Farrell, 1999] Sureshbabu, N. and Farrell, J. A. (1999). Wavelet-based system identification for nonlinear control. *IEEE Transactions on Automatic Control*, 44(2):412 – 417.

[Toliyat et al., 2003] Toliyat, H. A., Abbaszadeh, K., Rahimian, M. M., and Olson, L. E. (2003). Rail defect diagnosis using wavelet packet decomposition. *IEEE Transaction of Industry Application*, 39(5):1454 – 1461.

[Toshiba Corporation and NEC Corporation, 2002] Toshiba Corporation and NEC Corporation (2002). Toshiba and NEC Jointly Propose Next Generation High Capacity, Blue-laser DVD Format. http://www.nec.co.jp/press/en/0208/2901.html.

[Ulfarsson et al., 2002] Ulfarsson, M., Benediktsson, J., and Sveinsson, J. (2002). Wavelet feature extraction and genetic feature selection for multisource data. In *IGARSS '02. 2002 IEEE International Geoscience and Remote Sensing Symposium*, Toronto, Canada.

[van Helvoirt et al., 2004] van Helvoirt, J., Leenknegt, G. A. L., Steinbuch, M., and Goossens, H. J. (2004). Classifying disc defects in optical disc drives by using time-series clustering. In *Proceedings of the American Control Conference 2004*, Boston, MA, USA.

[Vetterli and Kovačević, 1995] Vetterli, M. and Kovačević, J. (1995). *Wavelets and subband coding*. Prentice Hall PTR, Saddle River, NJ, first edition.

[Vidal et al., 2001a] Vidal, E., Andersen, P., Stoustrup, J., and Pedersen, T. (2001a). A study on the surface defects of a compact disk. In *Proceedings of the 2001 IEEE International Conference on Control Applications*, Mexico City, Mexico.

[Vidal et al., 2001b] Vidal, E., Hansen, K., Andersen, R., Poulsen, K., Stoustrup, J., Andersen, P., and Pedersen, T. (2001b). Linear quadratic control with fault detection in compact disk players. In *Proceedings of the 2001 IEEE International Conference on Control Applications*, Mexico City, Mexico.

[Vidal et al., 1998] Vidal, E., Shøien, H., and Rasmussen, J. (1998). Focus control of the optical pick-up in a cd player. Master's thesis, Department of Control Engineering, Institute of Electronic systems, Aalborg University, Denmark.

[Vidal et al., 2001c] Vidal, E., Stoustrup, J., Andersen, P., Pedersen, T., and Mikkelsen, H. (2001c). Open and closed loop parametric system identification in compact disk players. In *ACC2001*, Arlington, Virginia.

[Vidal Sánchez, 2003] Vidal Sánchez, E. (2003). *Robust and Fault Tolerant control of CD-players*. PhD thesis, Department of Control Engineering, Aalborg University. ISBN: 87-90664-15-9.

[Warner Bros, 2002] Warner Bros (2002). Dvd production: Factions debate hd-dvd. Press news: http://www.medialinenews.com/issues/2002/september/dvd_10.shtml.

[Weerasooriya and Phan, 1995] Weerasooriya, S. and Phan, D. (1995). Discrete-time LQG/LTR design and modeling of a disk drive actuator tracking servo system. *IEEE Transactions on Industrial Electronics*, 42(3):240–247.

[Wesfreid and Wickerhauser, 1999] Wesfreid, E. and Wickerhauser, M. V. (1999). Vocal command signal segmentation and phoneme classification. In Ochoa, A. A., editor, *Proceedings of the II Artificial Intelligence Symposium at CIMAF 99*. Institute of Cybernetics, Mathematics and Physics (ICIMAF), Habana, Cuba.

[Wickerhauser, 1994a] Wickerhauser, M. (1994a). *Adapted Wavelet Analysis from Theory to Software*. A K Peters, Ltd., 1st edition.

[Wickerhauser, 1991] Wickerhauser, M. V. (1991). Fast approximate factor analysis. In Silbermann, M. J. and Tagare, H. D., editors, *Curves and Surfaces in Computer Vision and Graphics II*, volume 1610 of *SPIE Proceedings*, pages iii + 395, Boston. SPIE.

[Wickerhauser, 1994b] Wickerhauser, M. V. (1994b). Large-rank approximate principal component analysis with wavelets for signal feature discrimination and the inversion of complicated maps. *Journal of Chemical Information and Computer Science*, 34(5):1036–1046.

[Wisniewski and Stoustrup, 2001] Wisniewski, R. and Stoustrup, J. (2001). Generalized H_2 control synthesis for periodic systems. In *American Control Conference 2001*, Arlington, Virginia, USA.

[Wook Heo and Chung, 2002] Wook Heo, J. and Chung, J. (2002). Vibration and noise reduction of an optical disk drive by using a vibration absorber. *IEEE Transactions on Consumer Electronics*, 48(4):874 – 878.

[Xiangxun, 2002] Xiangxun, C. (2002). Wavelet-based detection, localization, quantification and classification of short duration power quality disturbances. In *IEEE Power Engineering Society Winter Meeting*, New York, NY, USA.

[Yao et al., 2001] Yao, L., Wang, A.-M., and Cheng, Y.-F. (2001). Track seeking hybrid fuzzy controller for the compact disc player. In *The 10th IEEE International Conference on Fuzzy Systems*, Melbourne, Australia.

[Ye et al., 2002] Ye, H., Wang, G., and Ding, S. X. (2002). Integrated design of fault detection system in time-frequency domain. *IEEE Transaction on Automatic Control*, 47(2):384–390.

[Ye et al., 2003] Ye, H., Wang, G., and Ding, S. X. (2003). A new fault detection approach based on parity relation and stationary wavelet transform. In *Proceedings of the American Control Conference 2003*, Denver, CO, USA.

[Ye et al., 2004] Ye, H., Wang, G., and Ding, S. X. (2004). A new parity space approach for fault detection based on stationary wavelet transform. *IEEE Transaction on Automatic Control*, 49(2):281–287.

[Yeh and Pan, 2000] Yeh, T.-J. and Pan, Y.-C. (2000). Modeling and identification of opto-mechanical coupling and backlash nonlinearity in optical disk drives. *IEEE Transactions on Consumer Electronics*, 46(1):105–115.

[Yen et al., 1992] Yen, J.-Y., Lin, C.-S., and Li, C-H. Chen, Y.-Y. (1992). Servo controller design for an optical disk drive using fuzzy control algorithm. In *IEEE International Conference on Fuzzy Systems*, San Diego, CA, USA.

[Yokoyama et al., 1994] Yokoyama, E., Nagasawa, M., and Katayama, T. (1994). A disturbance suppression control system for car-mounted and portable optical disk drives. *IEEE Transactions on Consumer Electronics*, 40(2):92–99.

[Zhou et al., 2002] Zhou, Y., Steinbuch, M., and Kostic, D. (2002). Estimator-based sliding mode control of an optical disc drive under shock and vibration. In *CCA02*, Glasgow, Scotland, UK.

[Zhu et al., 1997] Zhu, S., Choo, F., K.S., L., Chan, C., and Kong, P. (1997). Servo system control in digital video disc (dvd). In *ISCE '97 - Proceedings of 1997 IEEE International Symposium on Consumer Electronics*, Singapore, Singapore.

Appendix A

Electro-magnetic-mechanical model

Contenious time models The focus electro-magnetic-mechanical model is

$$\frac{e_{\mathrm{f}}(s)}{u_{\mathrm{f}}(s)} = 1.5497 \cdot 10^5 \cdot \frac{3.5869}{s^2 + 5.8741 \cdot s + 3.6148 \cdot 10^4} \tag{A.1}$$

The radial electro-magnetic-mechanical model is

$$\frac{e_{\mathrm{r}}(s)}{u_{\mathrm{r}}(s)} = 1.652 \cdot 10^6 \cdot \frac{0.6727}{s^2 + 5.8741 \cdot s + 9.7901 \cdot 10^4} \tag{A.2}$$

Extended discrite time models

$$\Phi = \begin{bmatrix} 0.9998 & -1.033 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.857 \cdot 10^{-5} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9978 & -0.03042 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.854 \cdot 10^{-5} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9998 & -2.7969 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.8569 \cdot 10^{-5} & 1.0000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.9978 & -0.0304 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2.8541 \cdot 10^{-5} & 1.0000 \end{bmatrix}, \tag{A.3}$$

$$\Gamma = \begin{bmatrix} 5.7138 \cdot 10^{-6} & 0 \\ 8.1628 \cdot 10^{-11} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 5.7137 \cdot 10^{-6} \\ 0 & 8.1628 \cdot 10^{-11} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{A.4}$$

$$\Psi = \begin{bmatrix} 5.7138 \cdot 10^{-7} & 0 \\ 8.1628 \cdot 10^{-12} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 5.7137 \cdot 10^{-7} \\ 0 & 8.1628 \cdot 10^{-12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{A.5}$$

$$\mathbf{C}_{\mathrm{CD}} = \begin{bmatrix} 0 & 2.2235 & 100.0000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2.2235 & 100.0000 & 0 \end{bmatrix}. \tag{A.6}$$

Appendix B

Optical model

In this appendix is the identified parameters in optical model listed. All the parameters are adressed with their respective the part functions in the model: $h_1, h_2, h_3, h_4, g_1, g_3, g_4$.

The identified parameters are not nescary the value of the physical parameters, since parameter identification do not find these physical parameters, but finds a set of parameters which given a model output close to the measured signals. As a consequence the physical units are left out in the tables A.1-A.4.

$l_{\mathrm{u},0}$	$l_{\mathrm{x},0}$	F	R	k_0
0.0010	$2.2698 \cdot 10^{-17}$	0.0129	$1.3604 \cdot 10^{-6}$	$1.3402 \cdot 10^{-14}$
P	r	r_0	b	α
1000	$6.2917 \cdot 10^{-11}$	$9.8656 \cdot 10^{-7}$	$1.2788 \cdot 10^{-6}$	$3.6100 \cdot 10^{-6}$

Table B.1: Identified parameters in $h_1(\cdot)$ and $h_2(\cdot)$.

r	r_0	b	α
$6.2917{\cdot}10^{-11}$	$4.96191{\cdot}10^{-7}$	$1.7936{\cdot}10^{-6}$	$2.2161{\cdot}10^{-6}$

Table B.2: Identified parameters in $h_3(\cdot)$ and $h_4(\cdot)$.

r	b_{s}	ρ_{pit}	ρ_{land}
$4{\cdot}10^{-7}$	$3.7621{\cdot}10^{-7}$	0.4291	1.6237

Table B.3: Identified parameters in $g_1(\cdot)$.

B.0.1 Polynomials

The coefficients in the polynomials in $h_3(\cdot)$ and $h_4(\cdot)$ are

$$p_{h_1} = \begin{bmatrix} -929.0194 \\ 599.1097 \\ 730.0290 \\ -468.6936 \\ -170.5094 \\ 124.2920 \\ 0.4108 \\ -12.6154 \\ 3.6062 \\ 0.5275 \\ 0.0326 \end{bmatrix}, \; p_{h_2} = \begin{bmatrix} -454.4608 \\ 209.0713 \\ 189.7289 \\ -107.0105 \\ 68.28776 \\ 1.9910 \\ -44.2038 \\ 4.8459 \\ 5.6990 \\ -0.5867 \\ 0.0480 \end{bmatrix}, \tag{B.1}$$

where the polynomials are defined as:

$$p_{S1}(x) = p_{h_1}(1) \cdot x^{10} + p_{h_1}(2) \cdot x^9 + \ldots + p_{h_1}(11), \tag{B.2}$$

$$p_{S2}(x) = p_{h_2}(1) \cdot x^{10} + p_{h_2}(2) \cdot x^9 + \ldots + p_{h_2}(11). \tag{B.3}$$

r	b_{s}	ρ_{pit}	ρ_{land}
$4{\cdot}10^{-7}$	$9.0686{\cdot}10^{-7}$	0.1094	0.4942

Table B.4: Identified parameters in $g_3(\cdot)$ and $g_4(\cdot)$.

Appendix C

Kalman estimators

The kalman gains are computed to be

$$M_\infty = \begin{bmatrix} 166.7491 & 0 \\ 0.0631 & 0 \\ -4.2311 \cdot 10^{-14} & 0 \\ 3.0313 \cdot 10^{-15} & 0 \\ 0 & 1.0432 \cdot 10^{3} \\ 0 & 0.1488 \\ 0 & 2.8887 \cdot 10^{-13} \\ 0 & -3.9815 \cdot 10^{-14} \end{bmatrix}, \tag{C.1}$$

$$L_\infty = \begin{bmatrix} 166.6535 & 0 \\ 0.0678 & 0 \\ -4.2312 \cdot 10^{-14} & 0 \\ 3.0301 \cdot 10^{-15} & 0 \\ 0 & 1.0426 \cdot 10^{3} \\ 0 & 0.1786 \\ 0 & 2.8946 \cdot 10^{-13} \\ 0 & -3.9807 \cdot 10^{-14} \end{bmatrix}. \tag{C.2}$$

Appendix D

Controllers

The normal focus PID-controller is

$$\frac{u_{\mathrm{f}}(z)}{e_{\mathrm{f}}(z)} = \frac{1476z^2 - 2630z + 1159}{z^3 - 1.332z^2 + 0.3598z - 0.02759} \tag{D.1}$$

The low-banded focus PID-controller is

$$\frac{u_{\mathrm{f}}(z)}{e_{\mathrm{f}}(z)} = \frac{1453z^2 - 2618z + 1165}{z^3 - 1.332z^2 + 0.3598z - 0.02759} \tag{D.2}$$

The normal radial PID-controller is

$$\frac{u_{\mathrm{f}}(z)}{e_{\mathrm{f}}(z)} = \frac{124.8z^2 - 222.3z + 97.97}{z^3 - 1.332z^2 + 0.3598z - 0.02759} \tag{D.3}$$

The low-banded radial PID-controller is

$$\frac{u_{\mathrm{f}}(z)}{e_{\mathrm{f}}(z)} = \frac{122.9z^2 - 221.3z + 98.48}{z^3 - 1.332z^2 + 0.3598z - 0.02759} \tag{D.4}$$